DOCUMENT RESUME

ED 040 850 SE 008 280

AUTHOR	Even, Alexander
TITLE	Patterns of Academic Achievement in Grade 12 Chemistry and Their Relationship to Personal, Attitudinal and Environmental Factors.
INSTITUTION	Toronto Univ. (Ontario).
PUB DATE	68
NOTE	421p.
EDRS PRICE	EDRS Price MF-$1.75 HC-$21.15
DESCRIPTORS	*Academic Achievement, *Chemistry, *Cognitive Tests, Evaluation, Foreign Countries, Grade Prediction, Science Tests, *Secondary School Science, Student Characteristics
IDENTIFIERS	Canada

ABSTRACT

Reported is a study designed to (1) describe the variations which occur in the attainment of cognitive objectives in high school chemistry, (2) identify patterns of achievement in terms of these cognitive objectives, and (3) investigate the relationship of achievement of these objectives and their patterns to certain personal, attitudinal, and environmental factors. The sample consisted of 2339 grade twelve chemistry students enrolled in the college-preparatory course in Ontario high schools. The criterion instrument was the Ontario Test of Achievement in Chemistry (OTAC), developed by the investigator to measure the cognitive objectives of knowledge, comprehension, application, and analysis. Other data gathered included (1) student's scores on the Scholastic Aptitude Test, Ontario edition (SATO), 1963-64, (2) students' background, and (3) final grades in chemistry. The principal findings were (1) that large variations in the attainment of the cognitive objectives of high school chemistry appear in the sample, (2) substantial correlations occur between OTAC scores and SATO Total Verbal and SATO Mathematics scores, as well between OTAC total scores and final grades in chemistry, and (3) the variables which make the most important contributions to the explainable variance of OTAC total scores are mathematics aptitude and verbal aptitude. (LC)

APR 1 1970

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

ED040850

PATTERNS OF ACADEMIC ACHIEVEMENT IN GRADE 12 CHEMISTRY AND THEIR RELATIONSHIP TO PERSONAL, ATTITUDINAL AND ENVIRONMENTAL FACTORS

by

ALEXANDER EVEN

A Thesis submitted in conformity with the
requirements for the Degree of
Doctor of Philosophy in the
University of Toronto
1968

"PERMISSION TO REPRODUCE THIS COPYRIGHTED MATERIAL HAS BEEN GRANTED BY Alexander Even

TO ERIC AND ORGANIZATIONS OPERATING UNDER AGREEMENTS WITH THE U.S. OFFICE OF EDUCATION. FURTHER REPRODUCTION OUTSIDE THE ERIC SYSTEM REQUIRES PERMISSION OF THE COPYRIGHT OWNER."

© Alexander Even 1968

SE 008 280

The author of this dissertation would welcome inquiries or other correspondence on matters relating to the field of inquiry of the study. Correspondence may be addressed to

Alexander Even,
The Ontario Institute for Studies in Education,
252 Bloor St.W., Toronto 5, Ontario, Canada,

OR

1375 Amber Crescent,
Oakville, Ontario, Canada.

(Current address will be maintained in the American Educational Research Association directory.)

ACKNOWLEDGMENTS

For many helpful suggestions the writer is indebted to the members of his supervisory committee: Dr. W.G. Fleming, chairman, Dr. L.D. McLean, and Dr. H.H. Russell. Thanks are due also to Dr. G.F. Atkinson, University of Waterloo, who acted as external examiner.

Dr. J.W. Burns, Professor of Chemistry, University of Western Ontario, and Dr. R.P. Graham, Dean of Science and Professor of Chemistry, McMaster University, critically read for scientific accuracy the prototype of the criterion test used in this study; N.J. Eisele, J. Ferris, and J.F. Tummon, all experienced teachers of high school chemistry, served as the panel of judges for classifying test items. To all these gentlemen the writer expresses many thanks.

I would like also to express deep appreciation to Dr. T. Bentley Edwards, Professor of Education, University of California at Berkeley, for the interest he has shown in the study and for his many helpful comments offered over the past years, both by correspondence and in person during my visit to Berkeley. Miss Betty Waterman, Assistant Specialist in Criminology, University of California at Berkeley, supervised the intricate processing required for the Guttman scaling program; to her I am most grateful. Dr. J.N. Morgan, of the Institute for Social Research, University of Michigan, was kind enough to interpret and discuss the results of the first run on the Automatic Interaction Detector program, and offered useful advice before and after processing; I acknowledge his assistance with thanks.

Appreciation is also expressed to the supervisors, principals, and teachers who endorsed and participated in the study. Without the cooperation of these colleagues the study would never have materialized.

Special thanks are due to Mrs. Barbara J. Finch, Assistant Editor, The Ontario Institute for Studies in Education, for her many hours of editorial assistance. Mrs. Maralyn Complin and Miss Gladys Didur patiently and expertly typed the manuscript.

I reserve particular tribute to my wife, Eleanor, for her moral support and assistance in ways too numerous to recount.

The program of studies which culminated in the present dissertation was made possible, in part, by financial awards from The Canada Council, the Ontario Secondary School Teachers' Federation, and The Canadian Education Association. The findings and recommendations in this report are not to be regarded as officially endorsed by these bodies.

TABLE OF CONTENTS

Page

ACKNOWLEDGMENTS ii

LIST OF TABLES x

LIST OF ILLUSTRATIONS xvii

I. INTRODUCTION 1

- Need for the Study 1
- Statement of the Problem 4
- Some Notes on the Organization of Secondary Education in Ontario 4

II. A REVIEW OF LITERATURE RELATED TO THE INVESTIGATION 11

The Objectives of Science Education 11

- Scientific Literacy 11
- Evolution of the Goals of Science Education 17
- Problems of Communication 21
- The Goals of Secondary School Chemistry 25
- The Gap Between Theory and Practice . 26

The Taxonomy of Educational Objectives ... 36

- Description 36
- Research prior to 1962 42
- Research subsequent to 1962 46
  - Test Guides, Examination of Taxonomy-type Tests, Test Banks, and Annotated Research Information Bulletins 46
  - Agreement of Judges in Classifying Test Items 50
  - Theoretical Considerations 54
  - Empirical Validation of the Taxonomy 57
- Other Schemes of Classification 66

Page

The Inventory of Choices 71

Rationale of the Inventory 72

Construction of the Inventory Scales 76

Results of Investigations Using the Inventory 80

Advantages of the Inventory 86

Criticisms of the Inventory 87

Summary 90

III. DESIGN OF THE STUDY 93

The Theoretical Framework of Pattern Analysis 94

Definition of Terms 94

The Function of Interaction in Profile Analysis 96

Prerequisite Treatment of the Data .. 98

Stabilizing of Scores 98

Additional Requirements for Comparison of Group Patterns ... 100

A Further Requirement in the Present Study 100

The Model, Formulas, and a Property of R_p 101

Procedure Used to Analyze Patterns .. 104

Factors Assumed to Influence Chemistry Achievement 106

Delimitation 108

General Hypotheses 109

Hypotheses Used in the Statistical Analysis 111

IV. CONDUCT OF THE STUDY 114

The Criterion Instrument 114

Experience with Published Tests 114

The Structure of OTAC 116

Page

Selection of the Sample 120

Data Collection Procedures 121

Preparation of the Data for Analysis 123

V. ANALYSIS OF THE DATA AND DISCUSSION OF FINDINGS 128

Characteristics of the Sample 128

Characteristics of OTAC 130

Conventional Test Statistics 130

Characteristics Pecular to Taxonomy-Type Tests 138

Subtest difficulty and Taxonomy. Category 138

Simplex Structure 139

Agreement of Judges 141

Pyramid Effect 143

AID Analysis 145

The AID Program 145

Program Restrictions 146

The Algorithm 146

Input Parameters 147

Types of Groups 148

Other Features of the Program .. 149

Importance and Significance 149

Interaction 150

Advantages of the AID Program .. 150

Method of Investigating Variables ... 151

AID Output 151

Results 152

Summary and Discussion of AID Analysis Results 172

Pattern Analysis 175

Page

Preliminary Treatment of the Data ... 175

Definitions of Overachievement, Underachievement and Normal Achievement ... 179

Methods of Forming Groups for Analysis ... 180

The Decision Process ... 181

The Program ... 181

Results ... 183

Groups Selected by the AID Program ... 183

Groups Stratified on Relative Achievement ... 186

Differences Between Means ... 201

Comparison of Findings with Those of Similar Studies ... 207

Taxonomy-type Tests ... 207

Factors Related to High School Chemistry Achievement ... 210

The Inventory of Choices ... 211

Summary and Discussion of Findings ... 213

VI. SUMMARY AND CONCLUSIONS ... 226

Summary ... 226

Conclusions ... 234

Principal Findings ... 234

Strengths of the Study ... 237

Limitations of the Study ... 238

Educational Implications of the Study 239

Suggestions for Further Research ... 241

APPENDIX A. The Ontario Grade 12 Course of Study in Chemistry ... 245

APPENDIX B. The Inventory of Choices Question Booklet and Answer Sheet ... 255

APPENDIX C. The Personal Information Questionnaire 260

Page

APPENDIX D. Materials Mailed to Participating Schools 262

APPENDIX E. The Development of OTAC 286

Preliminary Editions 287

Content Validation of 12 E 4 290

Improvements Incorporated in OTAC 293

APPENDIX F. The Ontario Test of Achievement in Chemistry (OTAC) 295

APPENDIX G. OTAC Item Analysis 303

Variability and Relative Dispersion 311

Skewness and Kurtosis 311

APPENDIX H. Descriptive Statistics 313

APPENDIX I. Pattern Analysis Tables 334

APPENDIX J. Equating of Scores on Different Editions of SATO 355

APPENDIX K. The Inventory of Choices - Reproducibility, Equivalence, and Stability of Scales; Migration 359

Reproducibility 361

Equivalence 362

Stability 362

Migration 363

Summary 365

APPENDIX L. Analyses of Covariance 367

APPENDIX M. Comparison of Teacher-Assigned Marks (Grades) and Objective Test Scores .. 375

APPENDIX N. Equations for Computing Residual Scores and Their Standard Errors of Measurement 382

Regression Equations 383

Page

Adjustment of Standard Error of Measurement 385

BIBLIOGRAPHY 388

LIST OF TABLES

Table Page

1. Synopsis of the Taxonomy of Educational Objectives (Cognitive Domain) 38

2. Analysis of Variance Model Used in Haggard's Method of Pattern Analysis 101

3. SATO Sample and Population Statistics 128

4. Main Item Statistics for OTAC 131

5. Projected Reliabilities of OTAC Subtests .. 134

6. Test-Retest Statistics for OTAC 136

7. Example of a Perfect Simplex 140

8. OTAC Subtest Intercorrelation Matrix 140

9. Extent of Inter-Judge Agreement in Allocating OTAC Items to Taxonomy Categories 143

10. Abbreviations for Independent Variables Used in Aid Analysis 153

11. AID Analysis for Run No. 1 - OTAC Total Score 156

12. AID Analysis for Run No. 1 - Contribution of Important Variables to OTAC Total Score Variance 157

13. AID Analysis for Run No. 2 - OTAC Total Score 159

14. AID Analysis for Run No. 2 - Contribution of Important Variables to OTAC Total Score Variance 160

15. AID Analysis for Run No. 3 - OTAC Total Score 163

16. AID Analysis for Run No. 3 - Contribution of Important Variables to OTAC Total Score Variance 164

Table Page

17. AID Analysis for Run No. 4 - OTAC Taxonomy Category 4.00 Score 167

18. AID Analysis for Run No. 4 - Contribution of Important Variables to OTAC Subtest 4 Score Variance 168

19. AID Analysis for Run No. 5 - Chemistry Marks 171

20. AID Analysis for Run No. 5 - Contribution of Important Variables to Chemistry Marks Variance 172

21. Correlations between OTAC and SATO Scores . 176

22. Comparison of Original and Adjusted Standard Errors of Measurement 179

23. Check Chart of Significant Profile Similarities for Groups Selected by the AID Program 184

24. Summary of Pattern Analyses for AID Group Ed Plans 1 185

25. Check Chart of Significant Profile Similarities for Theoretic-Immediate Groups 189

26. Check Chart of Significant Profile Similarities for Prudent-Theoretic Groups 190

27. Check Chart of Significant Profile Similarities for Immediate Educational Plans Groups 191

28. Summary of Pattern Analyses of Underachievers 192

29. Summary of Pattern Analyses of Overachievers 192

30. Summary of Pattern Analyses - Dichotomized Theoretic-Immediate Scale 193

31. Summary of Pattern Analyses of all Underachievers and all Overachievers Having a Theoretic-Immediate Score 194

Table Page

32. Summary of Pattern Analyses of All Underachievers and All Overachievers 195

33. Summary of Pattern Analysis of All Underachievers and Overachievers Combined 196

34. Summary of Pattern Analyses of Combinations of Underachievers and Overachievers Having the Same Theoretic-Immediate Score 199

35. Summary of Pattern Analyses of the Combination of Underachievers and Overachievers With Extreme Theoretic-Immediate Scores ... 200

36. Among-Subtests F's for Groups Having Congruent or Parallel Patterns Across Categories 2.00, 3.00, and 4.00 201

37. Differences Between Pairs of Means in Groups Having Congruent or Parallel Patterns Across Categories 2.00, 3.00, and 4.00 203

E-1. Administration Schedule of Preliminary Tests 287

E-2. Item Statistics for Test 12 E 2 288

E-3. Total Test and Subtest Difficulties of Test 12 E 4 289

E-4. Allocation of 12 E 4 Items to Subtests 290

E-5. 12 E 4 Main Item Statistics 291

E-6. Distribution of Item Response Positions in OTAC 294

G-1. OTAC Item Response Patterns 304

G-2. OTAC Item Statistics 306

G-3. OTAC Summary Statistics 310

H-1. Means and Standard Deviations of Continuous Student Variables 314

H-2. Medians, Means and Standard Deviations of Continuous Teacher Variables 315

Table Page

H-3. Means and Standard Deviations of Continuous School Variables Computed on Various Bases 319

H-4. Frequency Distributions, Means and Standard Deviations of Inventory of Choices 4-Point Scale Scores 320

H-5. Frequency Distributions, Means and Standard Deviations of Inventory of Choices 9-Point Scale Scores 321

H-6. Frequency Distributions, Means and Standard Deviations of Inventory of Choices 12-Point Scale Scores 323

H-7. Frequency Distributions of Categorical Student Variables 325

H-8. Frequency Distributions of Categorical Variables Classified by Teacher and by Student 328

H-9. Schools Participating in the Study 331

H-10. Summary of Schools Participating in the Study 333

I-1. Analysis of AID Group Ed Plans 1 Students Not Intending to Complete Grade 13 335

I-2. Analysis of AID Group Ed Plans 1 Students Intending to Complete Grade 13 335

I-3. Analysis of AID Group Ed Plans 1 All Students Combined 336

I-4. Analysis of Highly Immediate Underachievers Theoretic-Immediate Score of 0 337

I-5. Analysis of Moderately Immediate Underachievers Theoretic-Immediate Score of 1 .. 337

I-6. Analysis of Moderately Theoretic Underachievers Theoretic-Immediate Score of 2 .. 338

I-7. Analysis of Highly Theoretic Underachievers Theoretic-Immediate Score of 3 338

I-8. Analysis of Highly Immediate Overachievers Theoretic-Immediate Score of 0 339

Table Page

I-9. Analysis of Moderately Immediate Overachievers Theoretic-Immediate Score of 1 .. 339

I-10. Analysis of Moderately Theoretic Overachievers Theoretic-Immediate Score of 2 .. 340

I-11. Analysis of Highly Theoretic Overachievers Theoretic-Immediate Score of 3 340

I-12. Analysis of Immediate Underachievers Theoretic-Immediate Scores of 0 and 1 341

I-13. Analysis of Theoretic Underachievers Theoretic-Immediate Scores of 2 and 3 341

I-14. Analysis of Immediate Overachievers Theoretic-Immediate Scores of 0 and 1 342

I-15. Analysis of Theoretic Overachievers Theoretic-Immediate Scores of 2 and 3 342

I-16. Analysis of All Underachievers Having a Theoretic-Immediate Score 343

I-17. Analysis of All Overachievers Having a Theoretic-Immediate Score 343

I-18. Analysis of All Underachievers 344

I-19. Analysis of All Overachievers 344

I-20. Analysis of All Underachievers and Overachievers Combined 345

I-21. Analysis of All Underachievers and Overachievers Having a Theoretic-Immediate Score 345

I-22. Analysis of Highly Immediate Non-Normal Achievers Theoretic-Immediate Score of 0 .. 346

I-23. Analysis of Moderately Immediate Non-Normal Achievers Theoretic-Immediate Score of 1 346

I-24. Analysis of Moderately Theoretic Non-Normal Achievers Theoretic-Immediate Score of 2 .. 347

I-25. Analysis of Highly Theoretic Non-Normal Achievers Theoretic-Immediate Score of 3 .. 347

Table Page

I-26. Analysis of Non-Normal Achievers With Extreme Theoretic-Immediate Scores 348

I-27. Analysis of Highly Theoretic Overachievers Prudent-Theoretic Score of 0 349

I-28. Analysis of Moderately Theoretic Overachievers Prudent-Theoretic Score of 1 349

I-29. Analysis of Moderately Prudent Overachievers Prudent-Theoretic Score of 2 350

I-30. Analysis of Highly Prudent Overachievers Prudent-Theoretic Score of 3 350

I-31. Analysis of Theoretic Overachievers Prudent-Theoretic Scores of 0 and 1 351

I-32. Analysis of Prudent Overachievers Prudent-Theoretic Scores of 2 and 3 351

I-33. Analysis of All Overachievers Having a Prudent-Theoretic Score 352

I-34. Analysis of Highly Theoretic Normal Achievers 353

I-35. Analysis of Highly Theoretic Underachievers Profiles Across Categories 1.00, 3.00 and 4.00 353

I-36. Analysis of Highly Theoretic Underachievers Profiles Across Four Taxonomy Categories .. 354

I-37. Analysis of All Normal Achievers Profiles Across Four Taxonomy Categories 354

J-1. SATO Total Verbal Statistics - General Course 356

J-2. SATO Mathematics Statistics - General Course 357

K-1. Coefficients of Reproducibility 361

K-2. Equivalence of Three Scoring Methods 362

K-3. Stability (Test-Retest Correlation) Coefficients 363

Table Page

K-4. Migration of Orientation During the Grade 12 Year 364

L-1. Association Between Theoretic-Immediate Scores and OTAC Means Adjusted for SATO Total Verbal and Mathematics Scores 369

L-2. Association Between Prudent-Theoretic Scores and OTAC Means Adjusted for SATO Total Verbal and Mathematics Scores 369

L-3. Association Between Theoretic-Immediate Scores and Teacher-Assigned Marks (Grades) Means Adjusted for SATO Total Verbal and Mathematics Scores 370

L-4. Association Between Prudent-Theoretic Scores and Teacher-Assigned Marks (Grades) Means Adjusted for SATO Total Verbal and Mathematics Scores 370

L-5. Association Between Dichotomized Theoretic-Immediate Scores and OTAC Means Adjusted for SATO Total Verbal and Mathematics Scores 371

L-6. Association Between Dichotomized Prudent-Theoretic Scores and OTAC Means Adjusted for SATO Total Verbal and Mathematics Scores 372

L-7. Association Between Dichotomized Theoretic-Immediate Scores and Teacher-Assigned Marks (Grades) Means Adjusted for SATO Total Verbal and Mathematics Scores 373

L-8. Association Between Dichotomized Prudent-Theoretic Scores and Teacher-Assigned Marks (Grades) Means Adjusted for SATO Total Verbal and Mathematics Scores 374

M-1. Correlations Between Final Chemistry Mark, Final Average Mark, and OTAC Scores 376

M-2. Correlations of SATO Scores with Teacher-Assigned Marks and OTAC Scores 378

M-3. Contribution of Important Variables to OTAC Total Score Variance and Final Chemistry Mark Variance 380

LIST OF ILLUSTRATIONS

Figure Page

1. Two-way Classification of Interests according to Edwards and Wilson 75
2. Illustration of the Terms Profile and Pattern 94
3. Interaction as an Indicator of Profile Dissimilarity 97
4. AID Tree for OTAC Total Scores - Run No. 1 155
5. AID Tree for OTAC Total Scores - Run No. 2 158
6. AID Tree for OTAC Total Scores - Run No. 3 162
7. AID Tree for OTAC Taxonomy Category 4.00 Scores - Run No. 4 166
8. AID Tree for Chemistry Marks - Run No. 5 .. 170
9. Decision Process Flowchart 182
10. Patterns of AID Groups with Differing Immediate Educational Plans 187
11. Patterns of Overachievers and Underachievers 198
12. Patterns of Combinations of Underachievers and Overachievers with Extreme Theoretic-Immediate Scores 198

CHAPTER I

INTRODUCTION

Need for the Study

The general objectives of science education, as they appear in educational literature, are widely accepted. Carleton _et al_. (1960) have pointed out that these objectives have changed little in over forty years. Statements of the objectives of high school chemistry show no fundamental difference in concept, but are similar to the objectives of all science courses taught at the high school level.[1]

In view of the general agreement regarding the purposes of science teaching it is surprising to find that tests in high school chemistry typically are not designed to measure achievement of most of these objectives. When one examines commercially available tests one concludes that these devices test cognitive objectives almost exclusively. Moreover, undue emphasis seems to have been placed on recall or recognition of factual information; higher cognitive activities such as problem-solving and hypothesis-testing are represented much less frequently. Many chemistry tests claim to measure more than mere acquisition of facts, and many at the high school level do include items to measure understanding of concepts and principles, ability to interpret data, ability to draw sound conclusions from data, ability to recognize cause and effect relationships, and so forth. However, even these tests report achievement only in terms of a single score. The assumption on which this practice rests must be that the single score represents one homogeneous criterion variable. Some chemistry tests arrange items into appropriate subtests purporting to measure the higher cognitive objectives but fail to provide for subtest scores, thus neglecting to draw attention to the fact that different students obtaining the same total score may have done so in quite different ways. The validities of subtests purporting to measure achievement in the higher cognitive outcomes are seldom mentioned in standardized test manuals. The assessment of differential achievement in chemistry is therefore an area in which comparatively little research has been reported.

---

[1] Detailed consideration of the objectives is given in Chapter II.

Another area of evaluation which requires much more study is the influence of personal, attitudinal, and environmental factors upon the attainment of the objectives of a program of studies in high school chemistry. While many studies have reported on the relationship of chemistry achievement to nonintellective factors, again the overall score or grade predominates in the research design.

Anderson (1950, 1949) devised subtests to measure specific objectives of high school chemistry instruction and investigated the interrelationships of these subtest scores. Using the total score on the test as a criterion, Anderson studied the relationship of chemistry achievement to certain factors, primarily those of teacher qualification, teaching practices and teaching conditions in the schools. Attitudes of students were not a factor in this study, although their educational plans were considered.

Edwards and Wilson (1959b) used their Inventory of Choices to assign students to two attitude groups and studied gains in chemistry achievement made by each group. The groups were found to differ significantly in average gain. However, gains were measured only in terms of a total score; the possibility that different students could have made equivalent gains by being proficient in different areas of competence was not investigated in this study.

The two studies mentioned were concerned with achievement in "traditional" high school chemistry. The last decade has seen the rise of many new courses of study in high school science, of which the best known are the physics course of the Physical Science Study Committee (PSSC), the biology courses of the Biological Sciences Curriculum Study (BSCS), and the chemistry courses known as the Chemical Bond Approach (CBA) and the Chemical Education Material Study (CHEM Study). Two studies comparing differential achievement in traditional chemistry and in CHEM Study chemistry courses have been reported. June Anderson (1964) measured gains in specific cognitive objectives as did Herron (1965), but neither considered patterns of achievement, nor studied gains in relationship to nonintellective factors.

To date no study has been found in which patterns of achievement in chemistry have been studied and related to personal, attitudinal, and environmental factors: hence the need for a study such as the present one.

Statement of the Problem

The questions which this study attempts to answer are:

1. What variations occur in the attainment of cognitive objectives of high school chemistry?
2. What patterns of achievement occur with respect to these cognitive objectives?
3. What personal, attitudinal, and environmental factors are associated with achievement of cognitive objectives and patterns of achievement of these cognitive objectives?

Some Notes on the Organization of Secondary Education in Ontario

Under the British North America Act of 1867 autonomy in educational matters was granted to the provinces entering Confederation. Elementary and secondary schools are thus operated under the regulations of provincial Departments of Education, each headed by a Minister of Education who is an elected member of the Legislature. While local Boards of Education, elected by taxpayers, provide schools and hire teachers, the provincial Department of Education, among its other responsibilities, exercises considerable control over the content of courses of study, approves textbooks, and provides periodic inspection of the schools. Various provinces differ in the number of years of education provided in elementary and secondary schools: in Ontario elementary schooling extends from Grade 1 to Grade 8 and secondary schooling from Grade 9 to Grade 13.

Many changes are taking place in Ontario education at present, and those changes relevant to this discussion are included as footnotes. Additional remarks concerning changes in secondary school science courses will be found in the second last paragraph of this chapter. The educational situation described is that of 1964, the year in which the data for this study were collected.

An Ontario Secondary School Graduation Diploma of the General Course (formerly the Junior Matriculation Certificate) is awarded to successful candidates on completion of Grade 12, the final examinations being set and marked by the staff of the candidate's school. A

Secondary School Honour Graduation Diploma (formerly the Senior Matriculation Certificate) is awarded to the candidate who obtains standing in eight[2] papers at the Grade 13 level. In this case the examinations are external, and are set and marked by committees on which Ontario universities are well represented. These examinations are conducted annually in June.[3] However, the Secondary School Honour Graduation Diploma is in itself not a sufficient qualification for entrance to an Ontario university. Admission requirements vary with the university and also vary for different courses within the university. A common requirement for university entrance has been a standing of at least 60 percent in nine Grade 13 papers,[4] including English Literature and English Composition, and two papers in a second language. (The passing grade for all secondary school subjects at all grade levels is 50 percent of the marks assigned to the questions in each examination.)

The secondary school science program for the 2,339 Grade 12 students studied in this research was that prescribed for the General Course, a college-preparatory course. (Science programs in commercial and vocational courses are less rigorously prescribed.) The General Course science program consisted of the following sequence:

---

[2] Now reduced to seven credits including two credits in English.

[3] The Grade 13 Departmental Examinations were discontinued after June, 1967, and have been replaced by examinations set and marked in the candidate's school.

[4] Now seven papers. The two English papers have been combined into one, as have the two papers of each second language.

Grade 9 - General Science (physical science, mainly elementary physics)
Grade 10 - General Science (biology)
Grade 11 - Physics (traditional course)
Grade 12 - Chemistry (traditional course)

In Grade 13 these students could elect from chemistry, physics, botany, and zoology those subjects which might be required for entrance to the university course of their choice. At the Grade 13 level no minimum number of science courses is stipulated; many students might decide not to study any science subject, preferring to concentrate, say, in languages.

Some distinctive features of the science sequence offered to Ontario General Course students are the following:

1. All science courses are centrally prescribed; while schools may teach the topics in any course in any order, the topics listed must be taught. (Some courses list optional topics, but most topics are obligatory.)
2. Grade 9 Science is a compulsory subject.
3. After completing Grade 9 a student continues the regular progression outlined above unless he decides to discontinue the study of science,

4. A student rarely enrolls in Grade 11 Physics without successfully completing Grade 10 Science.
5. A Secondary School Graduation Diploma requires the completion of courses in Grade 11 and 12 English, History, and Physical Education, plus four options, one of which may be Science. To obtain credit for the science option the student must have completed both Grade 11 Physics and Grade 12 Chemistry.
6. Enrollment in Grade 12 Chemistry has been estimated to be 64 percent of the total Grade 12 enrollment, and 77 percent of the number of students completing the requirements for the Secondary School Graduation Diploma.[5] Since a course in high school chemistry is an entrance requirement for many schools of nursing, a substantial proportion of the enrollees are girls.
7. Only two textbooks (Cragg, Graham, and Young, 1959; Croal, Couke, and Louden, 1958) are approved for use in the Ontario Grade 12 Chemistry Course.
8. The Grade 12 Chemistry course is usually taught in five or six 40- or 35-minute periods per week. The Grade 13 course was designed to be taught in five 40-minute periods per week.
9. Since Grade 13 Chemistry is a second course in that subject, many topics that would be taught in a conventional one-year chemistry course are deferred

---

[5] Figures supplied by Dr. J. A. Keddy, Manager, Education Data Centre, Ontario Department of Education, Feb. 1, 1966.

until the second year and given treatment in greater depth. Equilibrium and reversible reactions, theory of atomic structure, ionization, normal solutions, detailed study of groups in the periodic table of elements, redox reactions, and organic chemistry are topics normally taught in Grade 13, although some teachers introduce some of these topics in Grade 12. Thus, the Ontario Grade 12 Chemistry course may not be as comprehensive as one-year chemistry courses offered in an eleven- or twelve-year elementary-secondary school program, but the two-year sequence of chemistry courses provided in Ontario's secondary schools is likely to be more comprehensive than most one-year chemistry courses offered elsewhere to college-bound students.

10. Agricultural Science, Parts I and II, is taught in some schools in place of Grade 11 Physics and Grade 12 Chemistry. There is a large overlapping of the subject matter between the Agricultural Science courses and Physics and Chemistry. Schools teaching Agricultural Science in place of Chemistry were not included in this study.

It should be noted that the Grade 12 students studied in this investigation formed the last group to follow the progression of science courses outlined. In September of 1961 a new course of study in Grade 9

General Science was introduced; this course consisted of elementary physical science (four to five months) and biology (mostly zoology) for the remainder of the year. In September of 1962 a new Grade 10 course in General Science was begun, and consisted of botany (approximately three months) and physical science for the balance of that year of study. Physics and chemistry continue to be taught in Grade 11 and Grade 12 respectively, but a revised, PSSC-oriented course in physics was introduced in September of 1963. A revised chemistry course for Grade 12 was introduced in September of 1967. Many of the topics formerly taught in Grade 11 and Grade 12 now appear in the new Grade 9 and Grade 10 courses. As a result the students succeeding the Grade 12 group studied in this research have a substantially different background in secondary school science.

A copy of the Ontario Grade 12 Course of Study in effect in 1964 forms Appendix A. A brief paper by Lucow (1965) provides some general information on Canadian education prevailing at the time the study began. For greater historical perspective the reader is referred to Phillips (1957). A recent work by Harris (1967) describes the Ontario educational system as it has evolved and at present exists; unique and distinctive features of the Ontario educational system are identified and explained.

CHAPTER II

A REVIEW OF LITERATURE RELATED TO THE INVESTIGATION

The purposes of science instruction have been abundantly discussed in the literature. It is the purpose of this chapter to review recent literature on the objectives of science education with particular reference to the secondary school teaching of chemistry. Since the Taxonomy of Educational Objectives and the Inventory of Choices of Edwards and Wilson play a central role in the investigation, the literature concerning these two devices is reviewed in considerable detail.

The Objectives of Science Education

Scientific Literacy

Johnson (1962) states that the overarching goal of science education is the development of a scientifically literate citizenry. A scientific literate exhibits curiosity about the how and why of materials and events and shows interest in the hearing and reading of those things which claim the time and attention of scientists. The interest is not lessened by unwelcome ideas and events. A scientifically literate person may not create scientific ideas, but he is conversant with current ideas in science.

Openmindedness about his own ideas and those of others also characterize the scientific literate. He is accurate in his observations and descriptions. His first expression of an idea is a hypothesis which is followed by further studies and critical observations; he is able to adjust his thinking in terms of new information. He expects the same accuracy in others and insists on being given the basis for the making of a judgment about the quality of ideas. This attitude carries over into all areas, such as philosophy, foreign affairs, and so on.

In Johnson's view scientific literacy is also a matter of feelings and values, but these must be founded on broad knowledge, which, in most cases, begins in the classroom. Thus the goals of science instruction in the schools are most important: one must consider the kinds of knowledge that should be presented and the specific attitudes and skills that should be instilled in students.

A major step in the study of scientific literacy was the formation, in 1965, of the Scientific Literacy Research Center at the University of Wisconsin. One of the first tasks of the Center was to define and describe the concept of scientific literacy; this was done by searching the literature published from 1946 to 1964 and analyzing the references to scientific literacy. Pella, O'Hearn, and Gale (1966a, 1966b) report on their analysis of more than 100 documents in the form of articles in professional journals, popular magazines, newsletters, conference reports and bulletins, and chapters in books. It is interesting to note that more than half of these documents were dated between 1960 and 1964.

Six referents to scientific literacy were isolated; these are listed here in order of frequency of mention. A scientifically literate person has an understanding of the

1) interrelationships of science and society
2) ethics that control the scientist in his work
3) nature of science
4) basic concepts of science
5) differences between science and technology
6) interrelationships of science and the humanities.

It should be noted that in the documents examined, the first three referents were mentioned at least twice as frequently as the last three referents.

In essence, the referents are concerned with what science is and what scientists do; these ideas are central to the whole discussion of the objectives of science teaching. Nagel (1963) presents a very concise account of the three aspects of contemporary science which help to define its nature and aims and to amplify the above referents:

1. Practical control over nature. This aspect of science is the ultimate justification to the majority of people, and it is this aspect which is emphasized to the neglect of the other aspects. Although this feature is not the sole or main motive of scientific inquiry, when such a motive is made focal, the picture of science is distorted, giving rise to the image of the scientist as an infallible miracle-worker. The tendency to make the scientific enterprise responsible for the barbarous uses to which its findings are sometimes put appears plausible when science and its technological fruits are equated.
2. Attainment of systematic but reliable knowledge. The aim of science is to demonstrate events and processes as instances of general laws and theories which formulate invariable patterns of relations between things. In pursuing this aim science satisfies the craving to know and understand, and thus has been a major force in the development of liberal civilization.

Science must be distinguished from "common sense," which has the limitations of being imprecise, fragmented, myopically utilitarian, and applicable only in routine experience; common sense beliefs are frequently mutually inconsistent, leading to the arbitrary adoption of one belief or another and a disregard for alternative possibilities for handling concrete problems. In fact science deliberately attempts to produce conclusions freed from the limitations of common sense.

Despite the general reliability of scientific findings neither scientific reports of specific matters of fact nor the theories and laws of science are infallibly true and in principle incorrigible. First principles of science remain corrigible.

3. The scientific method of inquiry. This is science's most permanent feature and is the ultimate warrant for confidence in the conclusions of scientific inquiry. The scientific method is common to all sciences—all sciences employ the same principles in evaluating the weight of evidence, the same canons for judging the adequacy of proposed explanations, and the same criteria for deciding between alternative hypotheses. Scientific method is thus the general logic employed for assessing the merits of an inquiry.

   Science is essentially a social institution, and its objectivity a product of a community of thinkers; scientific ideas must survive the cross fire of critical commentary that independently acting minds supply.

   Scientific inquiry does not consist just in the collection of facts; unless the facts are selected with reference to hypotheses, the inquiry is blind and aimless. Hypotheses are "free creations of the mind" analogous to creative effort in the arts. Hypotheses are tested against facts and against other hypotheses concordant with fact.

Perhaps the most important concept is that of controlled inquiry, wherein an eliminative procedure is instituted to ascertain the differential effects of a factor assumed to be relevant to the occurrence of a given phenomenon. The reliability of scientific conclusions is largely a function of the number and rigor of the controls imposed.

Measurement serves a threefold purpose in the conduct of inquiry: (a) to increase the precision of formulating facts and explanations so that the formulations can be tested more easily, (b) to make possible finer discriminations of traits which in turn enables statements of traits to be subjected to more rigorous controls, and (c) to permit more comprehensive comparisons between diverse events so that relations between things may be formulated accurately and systematically.

It is important to note that Nagel considers a scientific method to exist, but not for the purpose of obtaining new ideas or discovering solutions to problems. One must distinguish between the logic of verification and generalization vis-à-vis the creative construction of valuable hypotheses — a process that cannot be logically reconstructed.

Conant (1946, 1947), Bronowski (1951, 1955), Holton (1960, 1963), Shamos (1961, 1963), and Johnson (1962) may be taken as representative of the authors read in the analysis conducted by Pella, O'Hearn, and Gale (1966a). The points of view of these authors are essentially similar to those of Nagel whose exposition synopsized here was not included in their analysis. It is significant that none of the authors cited by Pella, O'Hearn, and Gale suggested all six referents to scientific literacy.

It would seem then that the purpose of science teaching is served largely by communicating to the student an understanding of the nature and aims of science.

Evolution of the Goals of Science Education

A brief summary of the historical development of the goals of science education should help in understanding the issues today confronting science educators in their effort to develop scientifically literate individuals. Johnson (1962) presents such a list. The goals are stated in their chronological order of development.

1. Mastery of subject matter. This aim has been the dominant aim of most science teaching. The result of

critical and careful studies by scientists, the body of scientific knowledge is growing and becoming more accurate through repeated checking; however scientific information is now estimated to be doubling itself within a period of 11-13 years (Strong and Benfey, 1960; Strong, 1959; Price, 1956). A major problem is to determine what knowledge should be selected for teaching purposes from the immense quantity now available. The appropriate grade level of the introduction of specific information is also a problem. Several decades ago it was realized that science instruction based largely on this goal is inadequate.

2. Instilling scientific attitudes. Some desirable viewpoints proposed were that strange and mysterious occurrences are explainable by natural causes; that final conclusions should not be based on one or few observations; that one should develop a continuing curiosity about materials and events. Unfortunately the development of scientific attitudes has been treated in haphazard fashion, often poorly spelled out and with few specific suggestions for implementation. Mastery of subject matter was still the dominant aim.
3. Great ideas or underlying principles of science. As lists of the underlying ideas and principles began to appear, a new focus for curriculum building was provided. Notable changes in approach ensued.

4. Personal creativity in scientific enterprise. The ability to identify and solve problems came to be regarded as an important aim of science. The emphasis was now placed on developing the ability to apply principles in new situations, to understand cause and effect relationships, and to select facts pertinent to a problem and draw sound conclusions from observed data.

In the development of the goals of science education the more recent goals have not superseded the earlier ones, but have incorporated these into the overall aim: the later goals are thus new dimensions added to the original goals still considered important.

The newer courses in secondary school science attempt to implement the integrated complex of goals. These new courses are alike in that they have reduced substantially the body of knowledge required in the course, and that they avoid rote memorization of a mass of information; in addition the stress on unifying principles has given them a coherence which former courses lack. Problem-solving and other creative acts are emphasized. However, the new courses do not emphasize uniformly the complex of goals discussed; differential emphasis on these goals is seen especially in the work of several national committees and in the new courses they have produced within the last decade.[1] A case in point in the field of secondary school chemistry is the divergent approaches of the Chemical Bond Approach group (structural approach) and the Chemical Education Materials Study group (laboratory approach), although neither group considers the course it has devised a definitive one. A similar case prevails in secondary school physics: the rigorous approach developed by the Physical Science Study Committee stresses the structure and evolution of physics; the program known as Harvard Project Physics emphasizes the humanistic background of the sciences, the effect which physics has had on other sciences, and the interaction of science and technology. (Harvard Project Physics Progress Report, 1967; Harvard Project Physics Newsletter 1, 1964)

Fischler (1963) and Rutledge (1962) discuss the consequences of the divergent paths now pursued in science education. Fischler points out that areas of duplication exist between the new courses in physics and chemistry, and between the chemistry courses and two of the new biology courses. Rutledge notes that two problems which have not been solved by the new emphases in high school are the matter of general versus specialized science education and the articulation of secondary school science courses with each other and particularly with science programs in the junior high school. In fact the new courses have intensified rather than resolved the problems; the integration of new courses into an overall program extending from elementary school to senior high school is beset with complications. Concern is also expressed that in the anxiety to stress the newer aims of science education, technology which is now an important part of our environment may be ignored. Rutledge stresses that the new courses must be constantly reevaluated; constant revision and redevelopment are essential. It is significant to note that most of the newer courses mentioned are at present undergoing or have just undergone substantial revision, based on the results of experimental teaching and searching criticism.

---

[1] For brief descriptions of these courses see Lockard (1967), Gatewood and Obourn (1963, pp.362-371) or Science Education News (December 1961). Interpretive summaries have been provided by Haney (1966, pp.3-18) and Hurd (1962).

Problems of Communication

To compare statements by various writers of the goals of science education is difficult, since the goals are expressed with varying specificity. The statements of objectives which have been mentioned thus far may be regarded as being so general that the teacher finds them of little or no use in carrying out classroom tasks or directing his efforts to better instruction. Burnett (1957, pp.175-176) feels that aims must be stated in considerable detail and related to desired behaviors so that they may be translated easily to the classroom situation. On the other hand it is possible for such aims to be so specifically defined and concomitantly numerous that they become unmanageable and obscure the purpose of science instruction. Illustrations of the proliferation possible are afforded by Martin (1948) who listed 300 principles of biological science significant for general education, and Wise (1941, 1942) who identified 272 principles of physical science significant for general education. These principles are, when specified in such detail, indistinguishable from specific aims.

Dressel (1960, p.60) feels that the proliferation of objectives causes difficulty and suggests that a few objectives of special significance should be clearly identified and then emphasized among science teachers. Both Burnett and Dressel are concerned with different aspects of the same problem, and while a broad statement of an objective must be broken down into detailed operational terms before it can be implemented, a few broad aims which can be readily kept in mind and from which detailed behavioral objectives can be deduced will provide focus to the teacher's thinking and practice. The six referents to scientific literacy listed by Pella, O'Hearn and Gale (1966a) serve this purpose.

The objectives of science education have therefore been stated in various ways in many publications such as textbooks on the teaching of science, reports and recommendations of curriculum study groups, courses of study, preambles to research studies in science education, surveys conducted by government and industry, standardized test manuals, articles and books dealing with science education and curriculum development, and so forth. While superficially the objectives may seem to vary widely and denote many different things, general agreement is found on critical examination, as was observed in the work of Pella, O'Hearn and Gale (1966b). The following works are representative of those containing recent statements of objectives: Blanc (1952); Burnett (1957, pp.19-24, 35-41); Dressel and Mayhew (1954, chap. 5); Fitzpatrick (1960, pp.5-7, 26-29, 51); Hurd _et al_. (1960, chap. II); Johnson (1962); _Modern High School Physics: A Recommended Course of Study_ (1959, pp.6, 66-67); Morris (1961, pp.94-99); _Science: An Interim Report of the Science Committee_ (1963, pp.1-9); Sears and Kessen (1964, pp.3-6); _Secondary Modern Science Teaching, Part 1_ (1954, chap. 1); _The Teaching of General Science_ (1950, chap. 3); _The Teaching of Science in Secondary Schools_ (1958, chap. II); Washton (1967, pp.34-45, 86-87, 134-136).

In no case do the stated objectives contradict the aims of science education mentioned previously nor do they add any new fundamental concept to those concepts listed. It is reasonable therefore to conclude that the broad aims of science education are widely accepted. Carleton et al. (1960) claim that for forty years there has been general agreement regarding the purposes of science teaching: authoritative pronouncements of the objectives of science education made periodically since 1920 have shown remarkable similarity. A comparison of the statements of objectives made in the publications referred to above bears out the constancy and general acceptance of the aims.

To recapitulate, the main goal of science education is to instill in the student an understanding of the nature of science and its basic concepts; the interrelationships of science and society, the humanities and technology; and the ethics of the scientist. It must be emphasized that this aim is accomplished only by having the student behave like a scientist and pursue the activities of science, within the limitations of the student's knowledge and abilities, in a large number of situations. It is essential that the student experience the joys and frustrations of the scientific enterprise first hand, through involvement in selected representative scientific exercises.

The Goals of Secondary School Chemistry

In general the objectives of secondary school chemistry teaching are the same as those of all science teaching (Sutman, 1965); specific differences occur largely in the means of achieving the ends, and are dictated by the nature of the subject matter and laboratory procedures that pertain specially to chemistry. This similarity is reflected in the fact that comparatively few references to the objectives of teaching chemistry are found in the literature. Most literature dealing with the objectives of science education conveys the impression that a science such as chemistry adds no general aim to those stated. This is to be expected since chemistry is by nature a fundamental science, depending little on such sciences as biology and earth science, for example, for conceptual contributions.

Some recent publications point up the essential similarity of the aims of teaching high school chemistry and the goals of science education in general: Fitzpatrick, (in Pierce 1960, pp. vii-viii); Klubertanz (1955); Montean, Cope and Royce (1963, pp.36-37); Morris (1961, pp.95-96, 164); Pierce (1960, pp.3-4); Sutman (1965, p.292); "The Reed College Conference on the Teaching of Chemistry" (1958); Uricheck (1967, p.6); White (1967, p.12).

The Gap Between Theory and Practice

The foregoing discussion has emphasized that statements of the objectives of science teaching as they appear in the literature have changed but little and are generally accepted. There seems to be agreement in what _should_ be taught, but in practice what _is_ taught may not help pupils reach some of these objectives.

The _Encyclopedia of Educational Research_ (1960, p.1220) makes reference to the gap between theory and practice. Among others who are concerned about the discrepancy existing between statements of objectives and classroom practices are Sutman (1965, pp.291-292), Fischler (1965, pp.402-403), and Carleton _et al_. (1960, pp.152-153). Studies by Beauchamp (1932) and Obourn (1950) show that the inconsistency has persisted for some decades. That the situation is not confined to the North American continent is evidenced by the criticisms leveled at chemistry education practices in English secondary schools (Brown, 1962, pp.593-5; _Science in Secondary Schools_, 1960, p.120) and in Australian secondary schools (Short, 1962, p.1). Short sums up the situation as follows:

> Despite the efforts of teachers and others concerned, thousands of students have spent time and energy in the study of chemistry without reaching any significant measure of understanding of the scientific process, without ever engaging in a real experiment and without discovering what the subject is really about.

In the past decade research evidence that the gap exists in the secondary schools has been offered by Aylesworth (1960, pp.372-373) who concluded that a selected group of science teachers did not appreciate the meaning of problem-solving and their role in the teaching of this process. Allen (1959, pp.38-41) in a study of attitudes held by high school seniors found evidence of misunderstanding and ignorance of the nature of science and the interaction of science and society; several misconceptions were revealed in the image students held of the scientist. Mead and Métraux (1957), in a survey of more than 130 high schools, found that a minority of students had a favorable image of the scientist while a majority had a distinctly unfavorable impression.

The new "alphabet" courses (PSSC, CBA, CHEM, BSCS, and so forth) may be considered as designed to fill, or at least narrow, the gap between the aims of science education as reported in the literature and prevailing classroom practices. Pella's (1967, p.354) analysis of these new courses reveals that they make no mention of either the relationship of science and technology or the social implications of science; he notes also that new courses and conventional courses alike give no attention to "science and the humanities," one of the six referents to scientific literacy.[2] He reports that many teachers teaching the new courses continue to have pupils memorize large numbers of facts with little or no attention given to concept development. Sutman (1965, p.291) claims that the new courses seem not to be improving the comprehensive high school's preparation of students for work in "pure" abstract science. Anderson (1964) and Herron (1965) both present evidence to suggest that students of lower ability gain more in ability to analyze elements, relationships, or organizational principles of a communication when enrolled in conventional chemistry classes than when enrolled in the CHEM Study program.

---

[2] Harvard Project Physics was not included in Pella's analysis.

Reasons have been advanced to account for teaching practices falling short of stated aims. Johnson (1962) claims that teachers in general do not know the processes of science and scientists in action; teachers need actual research activity to come to realize what scientific activity is. Fischler (1965, pp.402-403) asserts that the process of inquiry is misunderstood by many teachers, and that teachers may not appreciate the purpose of the course they are teaching. A study by Behnke (1959) which found disagreement between high school teachers and scientists on fifty statements pertaining to the nature of science and scientists in society supports the claims of Johnson and Fischler. Sutman (1965, p.291-292) feels that the vast majority of school age youngsters are unable to think abstractly to the extent required by the new courses in science. Ausubel (1965, pp.259-260, 263-264) feels that lately the process of inquiry has been emphasized at the expense of organized subject matter. Sutman (1966, pp.495-496) agrees that the process or discovery approach has been overemphasized and claims that as a result overall improved understanding on the part of high school graduates has not occurred. Rutherford (1964) maintains that not much progress toward the teaching of science as inquiry can be expected until teachers become well grounded in the history and philosophy of science and are able to view scientific inquiry as part of the content of science itself.

The discrepancies between theory and practice show up most plainly in the evaluation procedures employed to assess the achievement of students in science. There the absence of emphasis on objectives other than acquisition of knowledge and solving of routine numerical problems is especially noticeable.

Reasons for the wide variation between stated objectives of science teaching and those which persist in the classroom are often attributed to the fact that examinations and tests used in the schools, and in larger administrative units as well, test mainly factual knowledge. This point is made by Burnett (1957, pp.238-239) and by Morris (1961, pp.100-142). Morris found excessive stress placed on the recall of factual knowledge in Australian external chemistry examinations. Brown (1962, p.594) quotes the chemistry panel of the Science Masters' Association in England as saying "examinations should test what it is desirable to teach rather than that teaching should be directed to what is readily examinable." The British Ministry of Education Inspectors in Science in Secondary Schools (1960, p.71) claim that the tendency to concentrate on measuring memorized facts results from the relative ease of construction of questions for this purpose and the difficulty of testing achievement in the non-factual areas. Bebell (1962, pp.4-6) believes that there is greater emphasis upon that which is easier to measure than upon that which is important to measure; he cites the neglect of more elusive and important learnings such as critical thinking, creativity, and problem solving because of the difficulty in developing the measuring tools.

Dressel (1960, p.59) says:

> Since one of the axioms of measurement is that objectives not tested in examinations are not real objectives to students, it behooves every teacher to include items in examinations which measure accomplishment of all of the real objectives of a course.... Objectives involving scientific methods and attitudes will then become explicit goals of science instruction, both to teachers and to their students.

The authors of The Teaching of General Science (1950, p.117) make this statement:

> It is one of the dangers of any system of examinations that we may come to value only what we can test, and this peril should be ever present in the minds of both teacher and examiner.

Brown (1962, p.593) points out that the external examination syllabus becomes, in the practice of teachers, identical with the teaching syllabus. In Ontario the courses of study for Grade 13 subjects are set by the Department of Education and until recently the science course prescriptions have been characterized by lack of flexibility. The external examination syllabus has been the teaching syllabus for Grade 13 Chemistry. In the experience of the present investigator many teachers of Grade 12 Chemistry consider an important function of that course to be preparation of the student for the examination in Grade 13 Chemistry; these teachers therefore have emphasized in their teaching and testing the somewhat limited range of objectives tested in the external examination which terminates the Grade 13 Chemistry course.

In spite of the rather obvious need to have tests constructed to measure the accomplishment of a wide range of goals, such tests are generally not available to the teaching profession. The limitations of commercially published.tests have been recognized by Ahmann and Glock (1963, p.350), Nunnally (1959, p.270), and Thorndike and Hagen (1961, p.289).

The situation just described is regarded as serious, since evaluation is used to ascertain what changes a course produces in pupils, to determine how the course produces such changes and what parameters influence the effectiveness of the course. Cooley and Klopfer (1963, p.73) claim that the selection of appropriate evaluation instruments is probably the most crucial aspect of research and development in the areas of curriculum and methods. The need for curriculum evaluation is stressed by Novak (1963, p.6) who described the work thus far as "superficial at best." The Encyclopedia of Educational Research (1960, p.485) concludes its section on evaluation with the following words:

> In general the research needs in the field of educational evaluation continue to be (a) improved procedures for identifying the significant educational outcomes and translating them into observable student behaviors; (b) improved devices for appraising student behaviors — improved in the sense of being more valid, more reliable, or more administratively feasible; and (c) improved ways of integrating the results of these appraisals into a comprehensive evaluation of a student or a school program.

There have been, in the last decade, many attempts to communicate to the teaching profession the serious nature of the problem described and to suggest ways and means of improving the situation. Teaching for Critical Thinking in Chemistry (1958, pp. 22-31), published by the National Science Teachers Association, Problem Solving Through Science (1959) prepared by the Northern California Science Committee, and Problem Solving Methods in Science Teaching (Mills and Dean, 1960), are some of the publications designed to show teachers how objectives other than the acquisition of knowledge may be attained. Klopfer and Cooley (1963) report on a project designed to use materials drawn from the history of science to convey important ideas about science and scientists; these History of Science Cases materials were designed for use within existing courses in high school biology, chemistry, or physics. The method has proven effective in increasing student understanding of science and scientists, with little or no loss of achievement in content of the high school courses.

In the area of evaluation publications have stressed the feasibility of testing for objectives other than factual recall. Nelson (1958) provided a guide to test quality and gave many examples of suitable items. Monaghan (1960) indicated some ways in which objective test items may be developed to evaluate thinking ability, and provided many examples. In England, Examinations Bulletin No. 3 (1964) and Examinations Bulletin No. 8 (1965) encourage the use of examination questions requiring higher level mental processes. Butts (1964) designed a "tab" test to evaluate problem solving ability. Hedges (1966) published a book devoted entirely to testing and evaluation in science; the book showed in detail how to go about constructing items to measure the various objectives of science teaching. Nedelsky (1965) published a book on science teaching in which several chapters are devoted to testing; a very comprehensive list of possible objectives of a course in physical science is provided together with examples of items useful in assessing student achievement in these objectives. Analyses of Science Tests (1959), published by the National Science Teachers Association, critically surveyed standardized tests available for use in science classrooms; the report stressed the dearth of well-constructed and well-validated tests in science and the tests' emphasis on retention of specific facts.

Anderson (1950, 1949), in a study characterized by excellent statistical design, devised subtests to measure (a) acquisition of factual information in science, (b) understanding of the principles of science, (c) understanding and use of the scientific method, and (d) acquisition of scientific attitudes; one part of his investigation dealt with the correlations between these subtest scores, their correlation with IQ, and the contributions of the three other subtest scores to the explainable variance of objective (c). In another part of his study Anderson attempted to determine which factors in the teaching situation contribute to the achievement of the objectives of science instruction. Unfortunately Anderson combined the results of the subtests measuring the specified objectives into one overall score, and thereby reduced the usefulness of the study.

Perhaps the greatest impediment to effective evaluation has been the inability of teachers to translate the broadly stated objectives of science instruction into operational terms. The need for casting the statements in behavioral terms has been sounded by Burnett (1957, pp.175-177), Fischler (1963, p.350), Watson (1962, p.282) and others. Cooley and Klopfer (1963) illustrated how items may be written to specified objectives and refined through analysis in order to evaluate educational innovations.

In recent years Bloom's "Taxonomy of Educational Objectives" has received much attention as a system of classifying behavioral objectives; Morris (1961, pp.88-90) and Washton (1967, pp.86-87) have commented on the suitability of this system for classifying the objectives of science instruction, and Hedges (1966) uses the system throughout his book.

One advantage in using the Taxonomy of Educational Objectives is that the shortcomings of evaluation instruments and the variance between stated objectives and classroom practices are thrown into sharp focus. A number of studies attest to this fact: Lawrence (Cox and Unks, 1967); McGuire (1963a); Morris (1961); Scannell and Stellwagen (1960); Tyler and Okumu (1965). Since the Taxonomy of Educational Objectives is reviewed in detail in the following section of this chapter, the findings of these studies will be treated there.

The Taxonomy of Educational Objectives

Description

The idea for a taxonomic classification of educational objectives arose during an informal meeting of college examiners at the 1948 American Psychological Association Convention in Boston. Bloom and others, confronted with the task of providing a meaningful frame of reference to facilitate communication among educators and examiners, met annually on an informal basis to develop a symbolic system for classifying the objectives of education; their product was published as the Taxonomy of Educational Objectives Handbook I: Cognitive Domain (Bloom *et al*., 1956). The Handbook is a group product, representing the contributions of more than thirty specialists in testing. This version of the Handbook resulted from the suggestions and criticisms of several hundred readers to whom the preliminary edition was distributed for critical examination.

The Taxonomy of Educational Objectives is the result of logical rather than empirical investigation, and forms, "a grand index of all the variables which instructors and educational testers have suggested measuring for the purpose of evaluating instruction" (Cronbach, 1960, p.375).

The complete taxonomy consists of three domains: cognitive (recall or recognition of knowledge, development of intellectual abilities and skills), affective (interests, attitudes, values, appreciations, emotional sets or biases), and psychomotor (muscular or motor skills, manipulation of materials or objects, acts requiring neuromuscular coordination). The affective domain has been elaborated in Handbook II (Krathwohl, Bloom, and Masia, 1965) but the psychomotor domain remains to be developed. A short description of the cognitive domain and some suggestions for its use appear in Cronbach (1960, pp.374-380); a more detailed account of the Taxonomy of Educational Objectives and its potential uses is found in Lindvall _et al_., (1964, chap. 3). Handbook I (Bloom _et al_., 1956) contains the most complete and detailed description of the cognitive domain.

As outlined in Table 1, the Taxonomy of Educational Objectives, Cognitive Domain[3] has six major sections: Knowledge, Comprehension, Application, Analysis, Synthesis, and Evaluation. Passing from simple to more complex behaviors usually associated with thinking, the cognitive domain spans objectives from simple recall of factual material to highly original and creative ways of combining and synthesizing new ideas and materials.

---

[3] Hereinafter referred to as the _Taxonomy_.

TABLE 1

SYNOPSIS OF THE TAXONOMY OF EDUCATIONAL OBJECTIVES (COGNITIVE DOMAIN)

1.00 KNOWLEDGE Remembering something in a form very close to that in which it was originally encountered.

- 1.10 Knowledge of specifics. Recall of bits of concrete information.
    - 1.11 Knowledge of terminology.
    - 1.12 Knowledge of specific facts.
- 1.20 Knowledge of ways and means of dealing with specifics. Includes methods of inquiry, chronological sequences, standards of judgment, patterns of organization within a field.
    - 1.21 Knowledge of conventions, accepted usage, correct form and style, etc.
    - 1.22 Knowledge of trends and sequences.
    - 1.23 Knowledge of classifications and categories.
    - 1.24 Knowledge of methodology for investigating particular problems of phenomena.
- 1.30 Knowledge of the universals and abstractions in a field. Includes organization of ideas by means of theories.
    - 1.31 Knowledge of principles and generalizations.
    - 1.32 Knowledge of theories and structures (as a connected body of principles).

2.00 COMPREHENSION Understanding of material being communicated, without necessarily relating it to other material.

- 2.10 Translation from one set of symbols to another (i.e., to go beyond recall and restate the material).
- 2.20 Interpretation. Explanation or summarization of a communication (i.e., a reordering or new view of the material).
- 2.30 Extrapolation. Extension of trends beyond the given data to determine implications, consequences, corollaries, effects, etc.
  Interpolation is considered part of this process.

3.00 APPLICATION The use of abstractions in particular and concrete situations.

TABLE 1 — Continued

4.00	ANALYSIS	Breaking of a communication into its parts so that the organization of ideas is clear.
4.10	Analysis of elements. Identification of the elements included in a communication, e.g., recognizing unstated assumptions.	
4.20	Analysis of relationships, e.g., skill in comprehending the interrelationships among the ideas of a passage.	
4.30	Analysis of organizational principles, e.g., recognizing techniques used in persuasive materials, such as advertising, propaganda, etc.	
5.00	SYNTHESIS	Putting elements and parts into a whole.
5.10	Production of a unique communication.	
5.20	Production of a plan, or proposed set of operations.	
5.30	Derivation of a set of abstract relations.	
6.00	EVALUATION	Judging the value of material and methods for given purposes. Qualitative and quantitative judgments about the extent to which materials and methods satisfy criteria.
6.10	Judgments in terms of internal evidence, e.g., logical consistency, fallacies in arguments, etc.	
6.20	Judgments in terms of external criteria, e.g., evaluating material with reference to facts or criteria developed elsewhere.	

Note.—This table is taken, with minor modifications, from Bloom et al. (1956) and reproduced by permission of Longmans Canada Ltd.

The Taxonomy of Educational Objectives was organized as an educational-logical-psychological classification system, with major emphasis given to educational considerations: the boundaries between categories should be closely related to the distinctions made by teachers in planning curricula or in choosing learning situations. The system has a logical basis in that terms used are defined with precision and used consistently. The classification is consistent also with relevant and accepted psychological principles and theories (Bloom et al., 1956, p.6).

The guiding principles underlying the Taxonomy are:

1. Only intended behavior is classified; actual behavior may differ from that intended.
2. The major distinctions between categories have been made to reflect the distinctions teachers make among student behaviors.
3. The Taxonomy has been logically developed and made internally consistent.
4. The Taxonomy has been made consistent with present understanding of psychological phenomena.
5. The classification has been developed in a relatively neutral fashion, avoiding partiality to any one view of education.

The classification scheme is intended to be hierarchical, the higher categories representing more complex and abstract behaviors than the lower categories. The classification scheme is also conceived to be cumulative, the higher categories being built upon and including the lower categories. Thus in Table 1 the major sections can be seen to be listed in order of complexity: in general, a person must know or be able to recall something before he can comprehend it, and must comprehend it before he can apply it; a person must be able to analyze elements before he can analyze organization.

It should be noted that in the Taxonomy intended behaviors are defined operationally. For example, some teachers require that their students "really understand" while others may desire their students to "grasp the core or essence." The authors of the Taxonomy ask:

> Do they all mean the same thing? Specifically what does a student do who "really understands" which he does not do when he does not understand? (Bloom et al., 1956, p.1)

The authors believe that by reference to a set of standard classifications, such as the Taxonomy, teachers should be able to define such nebulous terms.

The Taxonomy is seen by its authors also to be a classification of the student behaviors which represent the intended outcomes of the educational process. They remark:

> It is assumed that essentially the same classes of behavior may be observed in the usual range of subject matter content, at different levels of education,...and in different schools. Thus a single set of classifications should be applicable in all these instances. (Bloom et al., 1956, p.12)

The concept of behaviors transcending subject matter content areas is quite often referred to in the literature as "generality of process over content."

When the Taxonomy is used to classify test items its distinctive structure gives rise to two difficulties:

1. A person's behavior can be assessed only in relation to his background of experience. Hence the classifi-

cation of any particular test item can be made only after the learning situations preceding the test are known or assumed.

2. The more complex behaviors are built on and include simpler behaviors. This structural concept gives rise to problems in the scoring of test items and also leads to difficulties in the statistical analysis of tests which include items belonging to the higher categories of the Taxonomy. The difficulties are considered in greater detail in later sections.

Research Prior to 1962

In the six years following publication of Handbook I few reports of investigations using the Taxonomy were published. The situation was appraised by Harris (1962, p.105) as follows:

> With respect to achievement testing, it appears that there are no new principles being proposed for developing achievement tests. Further,...no one is trying to define achievement except in the obvious manner of labelling content. Apparently the Taxonomy of Educational Objectives has had no influence during this period....it still is unfortunate that there is no active work reported that seeks to employ this scheme to limit and distinguish among definitions of achievement or to check out the relationships among types of achievement suggested by it. We are in the doldrums!

Stanley and Bolton (1957) report several related studies in which graduate students in a measurement class who had studied the Taxonomy for four weeks classified test items according to the Taxonomy subcategories. On half the items five or more classifiers indicated perfect agreement. Among eight judges the agreement was quite high when only the six major levels of the Taxonomy were used. Items that were classified were taken from Gerberich (1956) (227 items), and from two prospectuses of the Graduate Record Examinations. The distribution of items among all categories of the Taxonomy but one was roughly the same for items from each source: approximately 50% of the items in Category 1.00 (Knowledge) were from each source, about 20% in Category 2.00 (Comprehension), about 7% in Category 3.00 (Application), less than 1% in Category 5.00 (Synthesis), and about 7% in Category 6.00 (Evaluation); only in Category 4.00 (Analysis) were the proportions quite different, with the Gerberich items accounting for 6% and the Graduate Record items accounting for 17% of their respective distributions. About 12.5% of the Gerberich items were judged to be in the affective or psychomotor domain. Stanley and Bolton conclude that sufficient agreement in assigning items to Taxonomy categories warrants the regular analysis of tests using the Taxonomy; they also note the apparent "rote-knowledge bias of the typical curriculum at all levels of education."

Scannell and Stellwagen (1960) used the Taxonomy to classify statements of objectives and test items from final examinations submitted by high school chemistry teachers. The principal findings were that (a) over 50% of the stated objectives and 60% of the test items were directly concerned with accumulation of knowledge, (b) final examinations seldom required students to exhibit the more complex cognitive skills (understanding of various degrees), and (c) seldom was there a direct relationship between the levels of stated goals and the levels of behavior required on the examinations.

Two unpublished dissertations report research based on the Taxonomy. Schmadel (1960) constructed tests of Evaluation and Synthesis, as defined by the Taxonomy, in an investigation designed to study the relationship of creative thinking abilities to school achievement. Creative thinking abilities predicted more variance in Synthesis than did either mental age or achievement measures, and accounted for almost one-third of the variance in Evaluation. The usefulness of this study's results, however, is quite likely limited by the small number of items comprising the Synthesis and Evaluation tests.

Morris (1961) employed the Taxonomy in a survey of the external chemistry examinations of Australia. Morris classified secondary school-leaving examinations (nearly all of the essay type) set by all Australian states during the period 1877 to 1960. He found comparatively few questions that could be assigned to Taxonomy Categories 4.00 and 5.00, and no questions that could be classified in Category 6.00. Computations made on Morris' data by the present writer show that the mean percentages of questions falling in each level of the Taxonomy were Category 1.00—64%, Category 2.00—14%, Category 3.00—20%, Category 4.00—1%, and Category 5.00—2%. Median values for the five categories were 65%, 12%, 20%, 0%, 0%, respectively. As a result of his exhaustive investigation Morris recommends that 20% of chemistry examination questions be in Category 1.00, 30% in Category 2.00, and 50% in Categories 3.00 to 6.00.

In addition to the research described above, some references to the Taxonomy have appeared in the literature. Dressel and Nelson (1956) edited a folio of test items consisting of more than 800 pages of items in the biological and physical sciences. Most of these items were classified according to the Taxonomy by their contributors. In this folio there also appeared a discussion of the Taxonomy and many examples of science items whose taxonomic assignment was explained in detail. Nelson (1959) demonstrated how the Taxonomy could be used by teachers to construct a biology test and pointed out the advantages of constructing a test on such a framework. Dyer (1960) appealed for wider use of the Taxonomy in test construction, while Urdal (1960) drew attention to the appropriateness of the Taxonomy to secondary school science objectives and urged the systematic development of science education objectives and planned evaluation. Cronbach (1960) devoted several pages to an exposition of the Taxonomy and its suitability for classifying test items.

Research Subsequent to 1962

Following 1962 a number of studies concerned with the Taxonomy of Educational Objectives were reported. Those studies relevant to the present investigation are summarized here in five categories: (a) test guides, examination of tests, test banks, and annotated research information bulletins, (b) agreement of judges in classifying test items, (c) theoretical discussions of problems arising from use of the Taxonomy, (d) empirical validation of tests constructed on the framework of the Taxonomy, and (e) research using classification schemes similar in concept to the Taxonomy.

Test Guides, Examination of Taxonomy-type Tests, Test Banks, and Annotated Research Information Bulletins

Klinckmann (1963) presented the Biological Sciences Curriculum Study's adaptation of the Taxonomy as a means of determining whether BSCS tests actually incorporate BSCS aims; a difficulty was encountered in classifying test items when the relevant prior learning experience of the students was not known. In this study two BSCS tests and one Cooperative Biology Test were analyzed.

Lombard (1965) drew attention to the use of the Taxonomy as a functional guide for the construction of better classroom tests, and provided general examples of items in each category of the Taxonomy to serve as prototypes for many specific items in different subject matter areas.

Lombard did not consider the fine distinctions between the categories of the Taxonomy as being of major importance, but stressed instead the importance of avoiding undue emphasis on the lower categories of the Taxonomy or on any one specific category. He also drew attention to the range of item difficulty levels within each category.

In its Cooperative Science Tests Handbook the Educational Testing Service (1964, pp.30-39) classified items of the Cooperative Science Tests according to the Taxonomy.

The Alberta Department of Education issued a publication (Ayers et al., 1965) which simplified and adapted the Taxonomy for Grade Nine Science and indicated the intended proportion of items falling into each category on the forthcoming external Grade Nine Science examination. The emphasis given to the various levels of the Taxonomy was Category 1.00—40%, Category 2.00—30%, Category 3.00—20%, and Categories 4.00, 5.00, and 6.00—10%. While no examples of items from the examination could be given, many examples from all levels of the Taxonomy were provided using nonsecure items.

Tyler and Okumu (1965) used the Taxonomy to classify course materials and examinations in a teacher education program and found (a) a noticeable discrepancy between course descriptions and actual behaviors, and (b) a lack of attention given to developing non-Knowledge cognitive skills. They conclude that the Taxonomy provides a useful structure for looking at course behaviors.

Lawrence (Cox and Unks, 1967) reports on the results of classifying more than 4,500 items from 74 randomly selected social studies tests obtained from 63 high schools in southern California. Approximately 98% of the items were classified in Category 1.00 (Knowledge) and 75% of the total items fell in the one subcategory (1.12) Knowledge of Specific Facts.

An in-progress study by Grobman (Cox and Gordon, 1966) reports that the Taxonomy has been found a useful general guide for test construction and for workshops in writing Biological Sciences Curriculum Study tests. Grobman also reports the Taxonomy valuable for focusing attention on the kinds of objectives that should be tested; she mentions that difficulty has been experienced in writing multiple-choice-type items for the higher levels of the Taxonomy.

Smith (1968) reports on his efforts to construct scalable sets of items and gives an example of an item set in which satisfactory scalability was obtained. Smith's experience in attempting to construct sets of items in which each level of the _Taxonomy_ is represented in sequence in the same subject matter area has led him to believe that multiple-choice items in Synthesis, since they are deductive, are actually testing an analysis-type behavior; however this analysis-type behavior differs from the Analysis level of Bloom's Taxonomy as presently constituted. Smith also found Evaluation items very difficult to construct and no longer considers an Evaluation item necessary to complete a scalable set of items.

Lessinger (1963) describes how, in one California high school district, teachers were taught to use the _Taxonomy_ and encouraged to submit items classified on the basis of the _Taxonomy_ to district test banks to provide item pools from which classroom tests could be constructed.

Kellogg (1964) reports the development of an American History Test Bank as part of an in-service training program. Items constructed according to the _Taxonomy_ were contributed by secondary school teachers. Inspection of these items revealed that the higher the Taxonomy level the higher the proportion of free-response items, with Categories 5.00 and 6.00 composed exclusively of free-response-type items.

The establishment in England of a bank of test items in secondary school mathematics is reported by Wood (Cox, 1966).

The establishment of a *Taxonomy* clearing house at the University of Pittsburgh under the leadership of Dr. Richard C. Cox has provided a most useful facility for collecting and disseminating information concerning the *Taxonomy*. The most recent select and annotated bibliography (Cox and Unks, 1967) reports 62 studies and projects completed, papers read, published, or filed. The list of in-progress studies and utilization of the *Taxonomy* (Cox and Gordon, 1966) reports 22 studies and projects proposed or underway; a later list (Cox, 1966) contains 8 additional entries.

Agreement of Judges in Classifying Test Items

The original purpose of the *Taxonomy* was to facilitate communication among educators and examiners. The categories and subcategories should therefore "mean the same thing" to many different people. One may operationally define "mean the same thing" by stating that experienced teachers and test-constructors should be able to reach a consensus as to where in the *Taxonomy* an item should be classified, provided the judges are aware of the prior educational experiences of the particular group of pupils for whom the test is designed, assuming, of course, that prior educational experience is reasonably uniform for that group. A number of studies have been conducted to determine whether the terminology of the Taxonomy is sufficiently precise to permit close agreement on the part of item classifiers. The work of Stanley and Bolton (1957) in this regard has already been described.

McGuire (1963b) reports the use of a modified form of the Taxonomy to classify the items of the 1961 examinations produced by the National Board of Medical Examiners. Four panels of three members each were formed to rate the items. Each panel represented a medical specialty and the three panel members were expert in that specialty. Each panel rated only those items in its specialty. The results were the same for each of the four panels: all three judges in each panel agreed unanimously on 61% of the 683 items rated; two of the three judges on each panel agreed on 93% of the items.

Stoker and Kropp (1964) report that when five judges independently classified items in a test according to major levels of the Taxonomy, 11 of the 36 items were classified congruent with the categories whose processes the items were intended to evoke. Four of the five judges agreed on another 9 items in the same test. Of the 16 remaining items, 14 were agreed upon by three of the five judges. Another test was rated by a panel of four judges who did not rate the previous test. In this case 11 of the 36 items received unanimous agreement perfectly related to the categories the items were intended to evoke; three of four judges agreed on 16 other items and two of the four agreed on the remaining 9 items.

Kropp and Stoker (1966, pp.19-23) report the results of investigations made in 1962. A panel of five judges rated the Reading Test of the Metropolitan Achievement Tests, Advanced Battery. Four months later three members of the same panel and three new members formed a second panel and rated the items a second time. The modal classification of each item by the first panel was compared to the modal classification of the item by the second panel. Of the 44 items, 24 received the same modal classification both times. For five items the modal classifications made by the two sets of judges were in fairly good agreement. Five of the items received different modal classifications by the two sets of judges. The items of another part of the Metropolitan Achievement Tests, Advanced Battery (Arithmetic) were rated by seven judges. Of 45 items only 6 received identical ratings by six of the seven judges. No item received identical ratings from all seven judges.

In the 1962 studies intra-judge agreement was found to be greater than inter-judge agreement. Group discussions revealed that most of the disagreements were the result of the different assumptions about the prior experience of students for whom the test was designed. Generally, judges who had the most experience teaching students in the age group specified by the test tended to rate the items lower in Taxonomy level than did judges with less teaching experience. Kropp and Stoker conclude that within a set of preconceptions on the part of a judge, the Taxonomy provides a rather unambiguous guide for classifying the items.

Winter et al. (1965, pp.12-17) describe the problems encountered in reaching consensus in selecting items designed for a series of unit tests in secondary school chemistry. Difficulty was experienced in writing multiple-choice items in the categories Analysis, Synthesis, and Evaluation and in assigning items to these categories specifically. The difficulty was partially solved by simplifying the Taxonomy to four levels which they termed Recall, Comprehension, Application, and Higher Competencies. Only those items which were of appropriate difficulty and discriminating power and whose classification was agreed upon by all judges were retained for the tests. The authors admitted however that the desired distribution of item types (Recall—40%, Comprehension—20%, Application—20%, and Higher Competencies—20%) was compromised in the final construction of the tests.

Theoretical Considerations

The use of the *Taxonomy* in test construction is attended by serious theoretical and practical problems. Kropp, Stoker, and Bashaw (1966) present a discussion of the major problems of validating the *Taxonomy* and point out the dearth of empirical evidence on the validity of the constructs underlying the structure of the *Taxonomy*.

Critical problems are:

1. The choice of proper response measure, particularly the difficulties produced by the confusing of the product response (right answer) with the process response (behavior the item was intended to evoke);
2. Conditions under which the response measure is collected, with special reference to equalizing content knowledge among students;
3. The nature of the Knowledge level, with suggestions that this level may be at least two-dimensional, and complications arising from the fact that knowledge is required for successful performance at higher levels;
4. Statistical problems arising from the use of hierarchical data, which include the inapplicability

of usual true score theory to subtests which are dependent on other subtests, the effect that this dependency has on item statistics, and the need for redefining the item analysis sample at each Taxonomy level to avoid considerable unreliability being introduced into the item analyses of the higher level subtests.

Means of validating taxonomy-type tests both internally and externally were also suggested. Internal validation may be supported by evidence of an inverse relationship between the mean score of each subtest and its taxonomic level. Alternatively, if scatterplots are made for scores from pairs of Taxonomy levels, the presence of a roughly triangular distribution of data points would support the presumption that the Taxonomy is ordered hierarchically according to complexity. Since the nature of the Taxonomy leads one to expect higher correlation between adjacent levels than between more remote levels, the emergence of a simplex from the subtest intercorrelation matrix provides evidence of the hierarchical and cumulative nature of the taxonomy-type test. Three ways of externally validating taxonomy-type tests are provided by:

1. Administering such tests to students at various grade levels to determine whether mental processes represented are learned behaviors (indicated by increasing subtest scores in successive grades);
2. Relating taxonomy subtest scores to scores of other

tests of cognitive abilities, using these to define factor scores, and then formulating equations for the regression of each of the taxonomy subtests on the set of factor scores. The number of factors required to account for reliable taxonomy test variance should increase directly with the taxonomy level of the subtest;

3. Studying the relationship of high and low level cognitive abilities to intelligence and creativity measures. Such studies would add to existing evidence of validity, which is at present slight.

A technical study by Cox (1965) has shown that selecting items from an item pool biases the proportions of items in each Taxonomy category so that these proportions are no longer equal to the proportions originally present in the pool. This effect is due to the fact that average discrimination values differ for the items in each major category. Items in Categories 1.00 and 4.00 when selected on their discriminating power received less emphasis in the test than in the item pool, while for Category 2.00 the reverse was true. No such effect was observed with Category 3.00 items. Statistical selection of items was also found to operate differentially for male and female tryout groups. Cox recommends that (a) the most discriminating items should be selected from within a particular category rather than from a total item pool, (b) the tryout group should have essentially the same ratio of males to females as the groups for which the final form of the test is to be used.

Empirical Validation of the Taxonomy

The Taxonomy of Educational Objectives is the result of logical rather than empirical investigation. If one accepts the logical coherence of the _Taxonomy_ one encounters many unanswered questions such as:

1. Do the items in different categories measure different things?
2. Are the major categories hierarchical and cumulative?
3. Do the subcategories form hierarchies within the major categories of the _Taxonomy_ and are the subcategories cumulative?
4. Do tests constructed on taxonomic principles possess a factor structure which supports the concept of a taxonomy?
5. Do tests built on taxonomic principles show patterns of achievement which vary with different groups of individuals?

The research reviewed in this section has attempted to answer one or more of these questions by analyzing student responses to items in taxonomy-type tests. In this respect the research reviewed differs from that mentioned in previous sections, as those investigations were not concerned with actual responses of testees.

McFall (1964) constructed a 35-item test in general science divided into subtests A (Knowledge) and B (higher level cognitive processes). Low correlation (.41) between scores on subtests A and B was believed to indicate that the test had construct validity. IQ, scores on the Stanford Achievement Test (Total), Stanford Achievement Test (Science), and science grades considered individually showed significantly higher correlations with subtest A than with subtest B.

McGuire, using a modified version of the Taxonomy, tested students at the University of Illinois College of Medicine. Her exploratory investigation (McGuire, 1963a) suggests, (a) that certifying examinations in medicine currently employed measure chiefly the recall of isolated information, (b) that reliable examinations of more complex intellectual processes can be designed, and (c) that varied patterns of student behavior are revealed in examinations by the process approach. In connection with (a), it is interesting to note that the present writer observed similar findings in his inspection of certifying examinations given by the Faculty of Dentistry of the University of Toronto.

McGuire (1963b) found evidence of hierarchical and cumulative structure in medical examinations constructed according to her modified taxonomy: adjacent levels correlated more highly than widely separated levels. Like

McFall, McGuire believes that the generally low correlations among subtests indicate that the subtests were measuring different abilities. She also concludes that (a) process analysis yields reproducible results when applied to medical examinations, (b) medical examinations can reliably test the complex intellectual processes, (c) Synthesis and Evaluation should be interchanged in the hierarchy.

Several studies support the hierarchical and cumulative nature of the Taxonomy. Thomas (1965) used Guttman's simplex analysis[4] to test the cumulative hypothesis of the Taxonomy. Approximately 100 students were tested in basic physical and biological sciences with items representing the first three levels of the Taxonomy. The simplex obtained from the correlation matrix supported the cumulative nature of the Taxonomy in Categories 1.00 to 3.00. In addition, correlations between Taxonomy subtests and reasoning ability (measured by the Cornell Conditional-Reasoning Test and the Cornell Class-Reasoning Test) increased as Taxonomy level increased. However, as in McFall's study, the correlations between Taxonomy subtests and IQ (measured by the California Test of Mental Maturity) decreased as the Taxonomy level increased. No significant differences were found between mean intelligence scores for subjects whose level scores were consistent with the cumulative hypothesis and those whose level scores were inconsistent.

---

[4] Considered more fully in Chapter V.

Ayers (1966) factor analyzed a 40-item multiple-choice test in which the items had been classified according to the Taxonomy. In addition to providing support for the hierarchical structure of the Taxonomy, the study suggests the analysis of longer tests which make use of more Taxonomy categories.

The hierarchical nature of the Taxonomy is also supported by the findings of Stoker and Kropp (1964). Since categories bearing higher numbers are postulated to be more complex than categories having lower numbers, one would expect that the mean difficulty of subtests would increase as the level of complexity increases. This was found to be the case with two tests administered to 1,000 high school students. In one of the tests, however, Evaluation did not fit the pattern, suggesting that it may be misplaced in the hierarchy or that the items for it were poorly constructed. Simplex analysis of 20 correlation matrices indicated that in half the cases Categories 1.00 to 4.00 were ordered correctly, but that Categories 5.00 and 6.00 were repeatedly out of order.

Stoker and Kropp postulated a factor structure, but very little evidence for this structure was found. Group factors such as a comprehension factor, application factor, etc., failed to emerge from the factor matrices.

Smith (1965) investigated the scalability of the Knowledge and Comprehension sublevels of the Taxonomy.

Five educational psychology concepts were chosen and 11 items written to test each concept; one item was prepared for each of eight selected Knowledge subcategories and the three Comprehension subcategories. The test was administered to 341 educational psychology students who had been introduced to the content through lecture or textbook. The average difficulty level of the Knowledge and Comprehension subcategories did seem to conform to the hierarchy of difficulty suggested by the Taxonomy. However, the correlations between items were so low that little relationship was suggested between items concerned with the same psychological concept. Low correlations (equal to values expected by chance) indicated little relationship between items placed in the same Taxonomy subclass. The intercorrelation matrix of sublevel scores failed to form a simplex. Smith observed that in all subcategories except Extrapolation the items varied greatly in difficulty (in Extrapolation all items were very difficult); the Knowledge item difficulties varied with the discrimination required by foils or distractors, while the Comprehension item difficulties varied with all factors which affect reading. The results led Smith also to conclude that Extrapolation might better be placed in Application since the process seems to require recall of a principle, understanding of the principle, and then application of the principle to the constructed situation.

In this matter Bloom[5] concurs.

In a study growing out of the one just described, Smith and Paterson (1965) investigated the scalability of a logical progression of four selected subcategories of the *Taxonomy*. Five psychological principles were selected for testing, and four items were written for each principle, with one item in each of the categories Knowledge of Terminology (1.11), Knowledge of Principles (1.31), Interpretation (2.20), and Extrapolation (2.30). The test was given to 156 educational psychology students who had been introduced to all the information necessary to answer the items written at the Knowledge level. No simplex emerged from the intercorrelation matrix; the hypothesis of hierarchical structure was therefore not supported. Also significant was the lack of relationship between items dealing with levels of understanding of the same principle. Another result was that Smith and Paterson expressed doubts as to the feasibility of using multiple-choice items as a basis for either accepting or rejecting a hypothesis about the effectiveness of a learning experience.

In both the studies in which Smith participated, the low correlations found among items indicated that the correlations involved varied not just with the content and process included, but with all factors bearing on item difficulty: terminology, psychological distance between foils, specific determiners, and so on. It was pointed out that any attempt to test the scalability of the classes of the Taxonomy should hold constant all factors affecting difficulty, with the exception of the components of understanding in each question.

---

[5] Personal interview, Chicago, May 3, 1965.

A few studies have been reported in which the effectiveness of new curricula and teaching techniques has been gauged in terms of Taxonomy. Anderson (1964) investigated the effectiveness of the Chemical Education Materials Study curriculum versus traditional chemistry courses in achieving some objectives of the Taxonomy. A test based on Categories 1.00 to 4.00 of the Taxonomy was given to 638 students in seven Florida high schools at the beginning of the chemistry course and again after five months. No significant differences were found in mean gains of the two treatment groups except in one instance: low ability students in the conventional course performed higher on the Analysis subtest than did their counterparts in the CHEM Study program. Anderson also reported good interjudge agreement and found that factor analysis supported the imputed hierarchy of the Taxonomy over the first four categories. The factor analysis also suggested differences in the cognitive attack used by the different treatment groups: the greatest changes in factor structure occurred in Levels 1.00 and 2.00 of the

Taxonomy for traditional chemistry students, while the greatest changes took place in Levels 3.00 and 4.00 for the CHEM Study groups. Anderson's finding may be somewhat limited by the low reliabilities of the subtests; the consequent large standard errors of measurement may have obscured other differences between the two treatment groups.

A study similar in type to Anderson's was conducted by Herron (1965, 1966). A chemistry test based on all six major levels of the Taxonomy was given to students in four Chicago schools during the second week of September and again the following mid-May. Posttest means for each treatment group were calculated using pretest scores on the same test as a covariate. CHEM Study students scored significantly higher on Application than did students in the conventional course, regardless of ability. CHEM Study high ability students scored significantly higher on Analysis than did high ability students in the conventional course, but, as in Anderson's study, low ability conventional course students scored significantly higher on Analysis than did low ability CHEM Study students. No significant differences in means between the two treatment groups were found in the other Taxonomy categories. Gains in Categories 5.00 and 6.00 were quite small. Factor analyses indicated that conventional students rely more on lower level cognitive abilities than do CHEM Study students, but other factor analytic results could not be clearly interpreted. Herron also reports satisfactory inter-judge agreement in classifying items.

The Taxonomy has been used to evaluate large-group—small-group instruction in high school chemistry (Schmitt et al., 1966; Winter et al., 1965). It was found that categories other than recall were not without ambiguity; recall items were found to be easily distinguished from other types of items. On the whole it was felt that the practical application of the Taxonomy to produce meaningful results is no easy task. A finding peculiar to the study was the high intercorrelation (.80 and higher) of subtests, even between Category 1.00 and other levels. A common factor influential in determining performance on all four categories was therefore indicated. The investigators suggest that this common factor might be (a) recall itself (or memory) which forms an essential component of any of the higher cognitive abilities, (b) general scholastic aptitude, or (c) specific chemistry aptitude. In their comparison of large-group and the conventionally taught (small-group) students the investigators found significant differences in adjusted means in all four revised taxonomic categories; these differences were consistently in favor of the students in the large-group program.

Two studies using the Taxonomy to identify differential patterns of achievement are reported with essentially negative results. Milholland (1964) reports that factor analysis of a test in which items were classified according to the Taxonomy provided little evidence that the subtest scores represent the objectives they were designed to measure. Zinn (1964) found that the Taxonomy was not differentiable in the test behaviors of students, and suggests conservative interpretation of the results of such differential tests.

Other Schemes of Classification

Cronbach (1960, pp.374-375) has drawn attention to the fact that research on the nature of proficiency variables has been neglected. There have been proposed classification schemes other than the Taxonomy of Educational Objectives for the purpose of identifying and categorizing proficiency variables, although these have not received the same attention in the literature as has the Taxonomy.

Ebel (1953) has formulated a classification scheme consisting of six categories arranged in hierarchical order according to their relevancy to common teaching objectives. Ebel's system is unique in that the recommended percentage of test items that should fall in each category is stated. The categories are arranged in ascending order of value as follows:

1. Content Detail items are based on instructional materials or procedures rather than on instructional goals. (0%)
2. Vocabulary items can be answered correctly by one who knows the meaning of a particular term. (Less than 20%)
3. Factual items deal with specific details of information and are answered primarily on the basis of recall. (Less than 20%)
4. Generalization items deal with general descriptive statements and are answered primarily on the basis of recall. (More than 10%)
5. Explanation items require understanding, but may be answered on the basis of recall. (More than 10%)
6. Application items deal with uses of information and often require original thinking. (More than 10%)

Ebel claims that this objective item categorization may be done even if the classifier has very little competence in the subject matter involved.

A study reported by Cook (1960) in which he hypothesized that item discrimination indices and item difficulty indices would increase with increasing relevance category yielded results which failed to support either hypothesis. Mean item difficulties were significantly different, but showed no observable trend. The trend of the mean discrimination indices was opposite to that hypothesized, the factual items being most discriminating. Three reasons to explain the results observed were presented: (a) the variation in numbers of items in each category in the tests studied, (b) the inadequate orientation of test constructors and prior experience of students in objective testing, and (c) possible independence between relevance categories and item statistics. The generality of process over content was not demonstrated in this study.

Gerberich (1956) developed a system of classifying educational outcomes and presented 227 examples of test items drawn from a variety of subject matter fields to demonstrate how these outcomes could be measured. Several of the 13 major outcomes reside in the affective and psychomotor domains. As an alternative to the classification scheme of the Taxonomy of Educational Objectives, Gerberich's system deserves consideration.

In the _Taxonomy_, the cognitive abilities higher than Knowledge bear strong resemblance to what is commonly called "critical thinking." The Watson-Glaser Critical Thinking Appraisal (1951-52, 1964) for instance, purports to test five aspects of critical thinking: ability to draw sound inferences from a statement of facts, to recognize assumptions implied by a statement, to reason logically by deduction, to reason logically by interpretation, and to discriminate between strong and weak arguments. The counterparts of these five aspects are easily found in the _Taxonomy_ categories Comprehension to Evaluation. Herron (1965), Anderson (1964) and Charen (1963) used gains in Watson-Glaser test scores as one criterion of the effectiveness of an experimental versus conventional approach in high school chemistry instruction.

Rust (1960) analyzed the responses of over 500 students on three different tests of critical thinking and found only one weak general factor in each test. The analysis suggested that all items within a subtest do not measure the same skills or abilities. Rust also found evidence to support the idea that critical thinking involves a large number of unique abilities and items of knowledge.

Research has also been reported in which investigators have not developed a logical framework of the complexity of the *Taxonomy* but have used a much simpler structure.

Tyler (1934) showed that permanence of learning was greater for skills that now would be assigned to higher level cognitive categories of the *Taxonomy*. Factual information was found to be quickly forgotten. These findings are in line with modern reinforcement theory and the claims of the constructors of the *Taxonomy*. The educational logical-psychological basis therefore is supported by this study.

Smith, Tyler, *et al*. (1942, chap. 1) in a series of studies of college and high school teaching identified the purpose the schools claimed to hold, and drew up instructions for specifying the objectives of education and detailing the construction of appropriate evaluative techniques.

Prior to the development of the Taxonomy, Furst (1950) investigated the relationship between content and process in secondary school subjects. He administered 27 tests in several subject fields to two groups of students at the beginning of the eleventh grade and again late in the twelfth grade. The two groups were taught by methods differing in the extent of integration of courses and emphasis placed on the development of higher mental processes. Despite the differences in the two educational programs, the correlational patterns for the two groups were nearly alike. The most important finding was that tests dealing with the same subject matter area had higher intercorrelations than tests dealing with the same mental process. Furst thus found little evidence for the generality of process over content; no evidence emerged of general abilities (extending across fields of subject matter) to apply principles, to think critically, or to interpret data. His findings on critical thinking in this respect support those of Rust.

An investigation of the nature of proficiency variables in high school chemistry is reported by Porter and Anderson (1959). In this study specified abilities in chemistry were correlated with each other and with IQ. The specified abilities were (a) understanding of functional facts and concepts in chemistry, (b) understanding and application of functional principles, (c) understanding and application of elements of scientific method, and (d) ability to use the basic skills in chemistry.

Using analysis of variance Porter and Anderson found significant differences in total chemistry score among the three groups stratified according to IQ, with total chemistry score increasing as IQ increased. This relationship was not observed to the same degree with specific abilities, except for ability (c). No significant differences in ability (a) were found for the three IQ groups. The four subtests correlated equally well (about .40) with IQ. The four abilities were not highly correlated; subtest intercorrelations ranged from .65 to .38. Analysis of covariance and factor analysis were not used in the study and no attempt was made to examine the subtests for evidence of a cumulative hierarchical structure.

The Inventory of Choices

The Inventory of Choices was developed by T. Bentley Edwards and Alan B. Wilson of the University of California at Berkeley to measure the attitudinal orientation of high school students toward their environment.

Intelligence and previous achievement of the student are the most often used predictors of achievement in school, but it is well known that students of high intelligence are not uniformly successful in school work.

Similarly, marks (grades) are not necessarily stable throughout a student's school years; many students who have done well in high school do poorly in college, and vice versa. Successful students are generally those who are willing to work at their studies and who possess favorable attitudes toward the study of school subjects. A measure of student attitudinal orientation should therefore aid in predicting success in school. The Inventory of Choices was developed to investigate the relationship of attitudinal factors to academic success, and to student commitment to specialization in either the natural sciences or social and literary studies.

Rationale of the Inventory

The theoretical basis of the instrument has been fully presented (Edwards and Wilson, 1958c), and a complete description of the construction and validation of its scales has been reported (Edwards and Wilson, 1959a). Only a few broad principles will be mentioned here.

Formal education is expected to provide the student with knowledge and skills for intelligent decision-making, as well as with a favorable attitude toward the process of deliberation. A normal individual is considered to have developed an integrated preference system; ends which are incompatible (because of scarcity of time and effort, as well as of mutual exclusiveness) are ordered into a hierarchy of value which makes choice possible. The variation in deliberative or evaluative activity prior to making a decision can provide one dimension for analyzing preferences. This dimension can be conceived as a continuum ranging from an immediate affectivity for proximate ends to an analytical and evaluative consideration of alternative forms of action and their possible consequences and relationship to long-term goals (Edwards and Wilson, 1958c, pp.280-282).

Objects of interest are either social or nonsocial; most people are interested in both persons and things, and in the interrelationship of the two, but the extent to which one or the other is emphasized places individuals on a social—nonsocial continuum which affords another dimension for the analysis of preference systems (Edwards and Wilson, 1958c, p.282).

The interest model thus contains two dimensions, each of a bipolar nature: (1) a preference for ends suggested by deliberate, abstract considerations as opposed to a preference for immediate, proximate ends, and (2) a preference for social as opposed to nonsocial objects. Intersection of these two dimensions yields four vectors or "poles" which can be described as follows:

1. Prudent: deliberative analytic orientation toward the social environment.

Reflects upon alternative possibilities of social action.

Concerned with the long-term consequence of acts.

Renounces opportunities for the immediate gratification of proximate ends where this may conflict with more remote or general values.

Seeks to rationalize his environment by widening his scope of cognition, rather than narrowing or compartmentalizing it, and is thus motivated toward the behavioral sciences.

2. Theoretic: deliberative analytic orientation toward the nonsocial environment.

   Characterized by interest in natural science and the use of reason to apprehend the nature of things.

   Has intrinsic interest in the study, laboratory and research activities of natural scientists.

3. Aesthetic: characterized by an affective appreciation of things, and direct sensory perception of the nonsocial environment.

4. Immediate: preference for proximate social ends.

   Characterized by a dependence upon the esteem and sanctions of others.

   Shows interest in social recognition and recreation.

   Approaches social issues intuitively, conforming to proximate social pressures.

   In the extreme, largely controlled by external pressures, actions impulsively disorganized, and each decision compartmentalized according to the exigencies of the moment.

These abstractions are summarized in Figure 1.

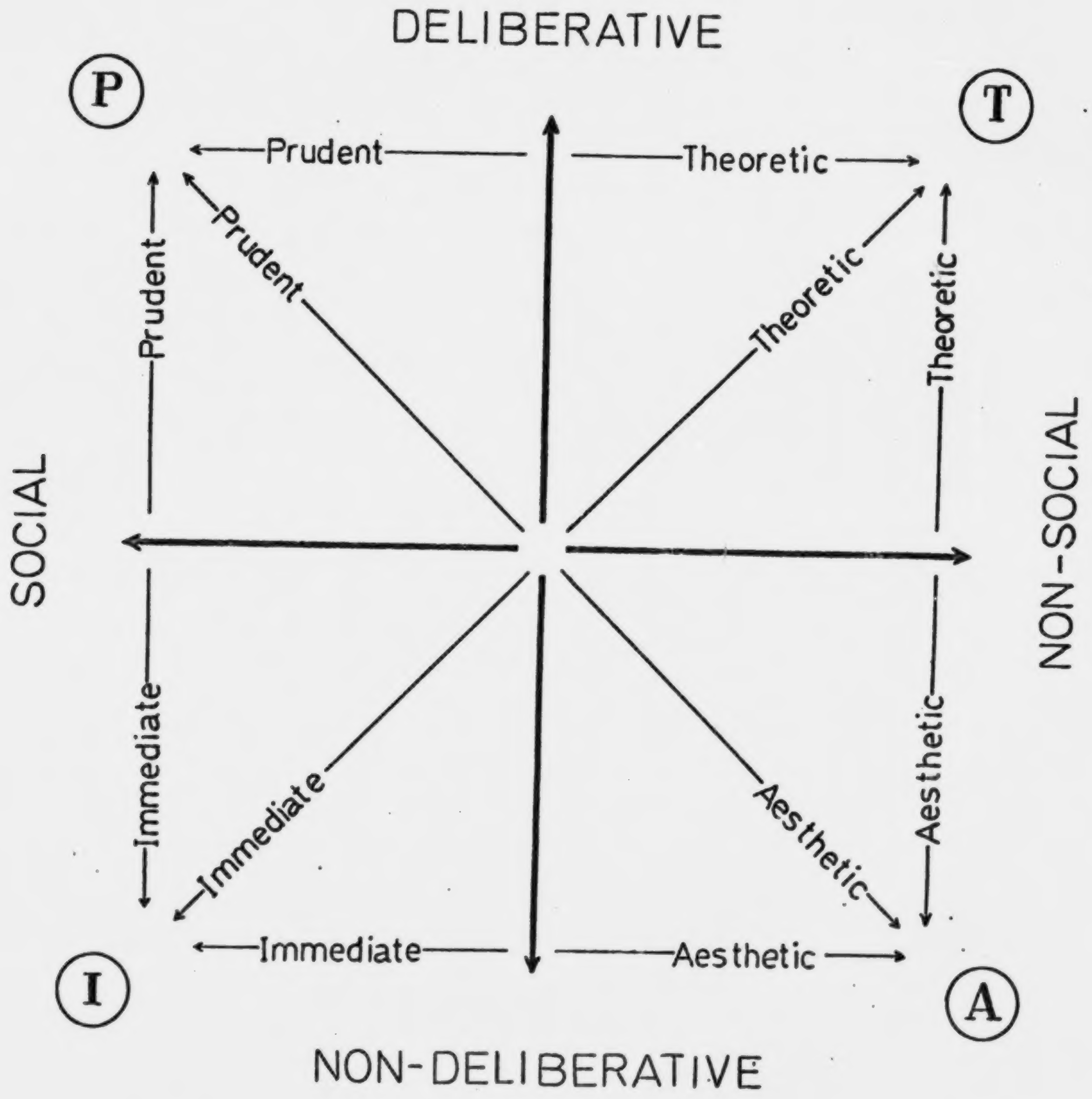

DELIBERATIVE
P
T
Prudent
Theoretic
Prudent
Prudent
Theoretic
Theoretic
SOCIAL
NON-SOCIAL
Immediate
Immediate
Aesthetic
Aesthetic
I
A
Immediate
Aesthetic
NON-DELIBERATIVE

Fig. 1--Two-way Classification of Interests according to Edwards and Wilson.

Note.—This figure is taken, with modification, from the article by T. Bentley Edwards and Alan B. Wilson published in the September 1959 Journal of Experimental Education, and reproduced by permission of Dembar Educational Research Services, Inc.

Pairing the four poles in all possible ways provides six scales for the assessment of the structure of attitudes; these six scales of the Inventory of Choices are labelled simply as Prudent-Theoretic (P-T), Prudent-Immediate (P-I), Prudent-Aesthetic (P-A), Theoretic-Immediate (T-I), Theoretic-Aesthetic (T-A), and Aesthetic-Immediate (A-I).

Construction of the Inventory Scales

A large number of short propositions were drafted by Edwards and Wilson, and from these, 180 were selected to form.the six scales. At all stages of development of the instrument, items were selected or judged adequate on the primary criterion of internal consistency, indicated by the extent to which the pattern of responses was cumulative. Other criteria adopted were those generally recommended in the literature on attitude scaling: relevance of content, clarity of meaning, appropriateness of vocabulary and content to the sample under study, and balance between positive and negative items to minimize response set (Edwards and Wilson, 1959a, pp.3-5).

A trial administration to 92 high school biology students and 50 undergraduate students in education, and a subsequent analysis of the items using the Guttman (1947) method resulted in a revised inventory of 72 randomized items, 12 items to each scale. Responses to items were indicated on a six-point Likert-type scale on which the level of agreement was designated as strong, moderate, or slight. This edition was administered to 325 high school students enrolled in college-preparatory physics and chemistry classes. The "H-technique" of Stouffer was then employed to form each scale into three "compound items," each containing four items (statements). With the H-technique each element in a compound item is as similar as possible to the other elements in that compound item, while each compound item is as different as possible from the other compound items in that scale. (Here "different" and "similar" are determined by the proportion of responses accorded each item.) An individual thus receives a score of 3, 2, 1, or 0 on each scale, according to the number of compound items responded to in a positive manner. To achieve a "positive" score on a compound item an individual must give favorable responses to two-thirds or more of the items forming that compound item. This reduction in sensitivity of the instrument increases its power to discriminate (Stouffer, 1952).

Scales constructed on the Guttman principle must show a cumulative response pattern within which items are ranked in order of popularity of response. Thus, in theory, given an individual's score on a scale, one should be able to reconstruct the pattern of responses made by that individual on that scale. This property of reproducibility, in Guttman's view, is characteristic of a genuine (that is, highly homogeneous) scale (Guttman, 1944; Guilford, 1954, pp.460-461). In practice, perfect reproducibility is seldom encountered, since the idiosyncrasies of respondents add a certain amount of disorder to the patterns of the responses; also the scales are rarely completely pure. However, a "coefficient of reproducibility" can be calculated for each scale, and a limit is set (usually .90) below which this coefficient should not fall if the scale is to be retained in the instrument. An example may clarify this point: If a scale has been found to have a coefficient of reproducibility of .92, this would mean that, given a respondent's score on that scale, one can reproduce his replies to the items of the scale with 92% accuracy (by predicting the cumulative pattern of the responses). In the larger high school sample mentioned earlier coefficients of reproducibility ranged from .90 to .95; no further revision of items was indicated.

The scales of the Inventory are therefore believed to have internal validity, based on their cumulative structure plus the grouping of items by means of the H-technique which eliminates most of the discrimination due to idiosyncratic factors. According to Guttman (1950, pp.305-311) the coefficient of reproducibility sets a lower bound to the reliability of the measures. All the measures described when limited by the Stouffer H-technique reached or exceeded a level of reproducibility of .87 in Edwards' and Wilson's most extensive study (Edwards and Wilson, 1958a, pp.35-40). The same estimate of unidimensionality provides evidence of validity.

Factor analysis has shown the dimensions of the Inventory of Choices to be relatively pure, with no second factor of consequence. The third factor brought out a cluster of "arty" or "highbrow" preferences (3 items) and the fourth factor a preference for solitary over social activities (3 items).[6]

The scales have been validated externally by several means: over 1,200 of the 3,750 students in the authors' most comprehensive survey (Edwards and Wilson, 1961) were interviewed individually, and the assessments made were compared with the scores of these students on the Inventory. Teacher ratings were obtained for all subjects in the study, and a survey of the subjects' hobbies and use of recreational time was carried out. In addition, a scientific demonstration was shown to a subsample of the students and an analysis of their individual attempts to solve the problem arising from this demonstration was made and compared to their Inventory scores.

---

[6] For a full discussion of this point see Appendix C of the final report made by Edwards and Wilson to the U.S. Office of Education (Edwards and Wilson, 1961).

Later administrations of the Inventory[7] to 894 British secondary school students and 979 eighth-grade students in Berkeley, California produced no discordant findings. The same is true for the results of an elementary school form administered to 319 third-grade students, 226 fifth-grade students, and 816 sixth-grade students in Berkeley. All methods of validation gave substantial support to the general premises on which the instrument was built.

Results of Investigations Using the Inventory

The central task of the Inventory of Choices was to identify salient differences in outlook and interest among students; these differences were expected to be relevant to their school achievement. In a pilot study (Edwards and Wilson, 1958b) the Inventory of Choices was administered to 325 students enrolled in physics and chemistry in one high school. It was found that the boys tended to have Theoretic interests while the girls had predominantly Prudent interests. Almost twice as many highly Prudent students received median grades of A, B+ or B as did highly Theoretic students (Edwards and Wilson, 1958b, p.186), but almost four times as many highly Theoretic students as highly Prudent students received high scores in a standardized chemistry test. However, the chemistry grades assigned to these same students by their teachers did not follow the pattern of the chemistry scores: students receiving any given grade from instructors were split almost equally between Prudent and Theoretic, with the Prudent students having a slight advantage (pp.189-190).

---

[7] Slightly modified to conform with British usage and idiom.

Scores on the Theoretic-Immediate scale revealed that boys had higher Theoretic interests than girls. Boys who had Theoretic rather than Immediate interests received higher grades in science and mathematics whereas the girls' Theoretic versus Immediate interests had no bearing on their science and mathematics grades (pp.193-194).

As part of the pilot study the gains in chemistry achievement during a two-semester period were analyzed. Achievement was measured by scores obtained on two forms of the Anderson Chemistry Test, one form being used as the pretest and the other form as the posttest. The subjects studied were 177 Introductory Chemistry students. When IQ, sex of student, and teacher were held constant by the techniques of analysis of covariance, Prudent students were found to have made, on the average, only 82% of the gain of Theoretic students (Edwards and Wilson, 1959b).

Subsequent to the pilot study the Inventory of Choices was administered to a sample of 3,750 students representative of the high school population in the San Francisco Bay area. In this study the findings of the pilot study were generally supported (Edwards and Wilson, 1958a, pp.53-57). The data collected in this study were further analyzed and incorporated into a more comprehensive investigation (Edwards and Wilson, 1961) in which an elementary school form of the instrument was developed and administered.

One of the findings of the more extensive study was that the Prudent-Immediate scale proved to be significantly related to achievement: 58% of the students ranked as most Prudent received median grades of A or B while only 18% of those ranked as most Immediate had comparatively high grades (Edwards and Wilson, 1961, p.C-21).

In the same study a surprising result was obtained when an elementary school version of the Inventory of Choices was administered. Prudent students in Grade 6 received a higher proportion of good grades than did Immediate students. Among Grade 3 students this pattern was reversed, the Immediate students receiving higher marks. When these students were classified along the Prudent-Theoretic dimension an analogous discontinuity was observed: in Grade 6 the Prudent students again received higher grades, while in Grade 3 the Theoretic students were the higher achievers. In this latter case the differences were not as pronounced as with the Prudent-Immediate scale, but they were still substantial.

These observations suggest that (1) requirements for early success in school are responsiveness and docility, while deliberation only later plays a significant role, and (2) beyond the primary grades, school programs fail to distinguish between students with Theoretic orientations and those who are essentially non-deliberative (Edwards and Wilson, 1961, pp.100-105).

A recent study by Maykovich (1966) compared scores on the Inventory of Choices with grades obtained by a sample of 61 boys over their four high school years. Inventory scores were available for the boys at the time of their high school entrance; the investigator administered the Inventory to the group for a second time at the beginning of the senior year. All attitude groups were homogeneous in ability and achievement at the start of the study.

The findings of Maykovich which are relevant to this discussion were:

1. There was a massive shift toward the Prudent pole on all scales in which the Prudent orientation was represented. A marked shift away from the Theoretic orientation took place on the Theoretic-Immediate and Prudent-Theoretic scales, which have a social orientation as the alternative.
2. Throughout the four years Prudent students were remarkably superior to all other groups in all subjects,

including the sciences and mathematics (except chemistry and geometry).

3. The relationships between attitudes and academic performance were of the same kind at twelfth-grade entrance as at ninth-grade entrance but with less prominent differences because of migration to the Prudent pole. This migration resulted in poorer students being included in the Prudent group and a consequent lowering of the average grade in that group. The movement away from the Theoretic pole resulted in the remaining students making better grades in mathematics and science in the twelfth grade than the Theoretic group had made in the ninth grade.
4. Attitudes held by the students in the twelfth grade are less important with respect to achievement than attitudes held in the ninth grade. No appreciable improvement in performance took place in the last two years of high school, regardless of any change of attitudinal orientation in any direction.

As a result of these findings, Maykovich concluded that:

1. The attitudinal orientation of students at high school entrance was vital to academic success.
2. Changes in attitude, even of a drastic nature, were not helpful. Performance patterns were set in the first two years.

3. Without Prudent skills, a student had little hope of academic success in high school.
4. Theoretic skills were of little avail in obtaining high grades; the student did as well without these skills as with them, even in the Theoretic subjects.
5. While the sample studied was not very large, the high levels of significance attained were so extreme that the need for further research was strongly indicated.

It should be noted that Maykovich used grades and grade point averages as measures of academic performance. However, as Edwards and Wilson have shown (Edwards and Wilson, 1958b, pp.189-190), the relationship between Inventory scores and grades is quite different from that between Inventory scores and standardized test results.

The findings of Maykovich are particularly relevant to the findings of the present study and will be referred to again in Chapter V.

To date, therefore, research employing the Inventory of Choices suggests a tendency for elementary and secondary schools to set standards of acceptance which cluster around the Prudent pole. Prudent students have been found to be higher achievers, overall, than Theoretic students, when measured by teacher-assigned grades. Prudent students therefore enjoy a strategic advantage in the competition for college admission. The fact that success in high school correlates with success in life generally could mean that society shows a more ready acceptance for the Prudent person than for the Theoretic. But the rejection of the Theoretic person may not be in the best interests of society. Although colleges may unwittingly place a premium on Prudent attributes in their admission procedures, the authors of the Inventory of Choices suspect that the academic role in nearly all graduate departments of universities is heavily loaded with Theoretic attributes. If this is true, a discontinuity in academic expectations occurs somewhere in the undergraduate years. This discontinuity must be highly frustrating to those with Prudent interests, but must come as a welcome change of pace to those with Theoretic interests who have survived.

Advantages of the Inventory

The present writer has chosen the Inventory of Choices as an attitude-measuring instrument in this study for several reasons. The underlying theory is attractive and comprehended without great difficulty. Teachers not having an extensive background in psychology and personality theory would be able to appreciate findings expressed in terms of the scales. The schema is particularly relevant to the understanding of attitudes of high school students toward their subjects. Students do not appear to consider the questions intimate or objectionable. The authors of the Inventory report[8] that almost no evidence of faking has been found, since the intent of the instrument is difficult to determine unless the respondent is conversant with the theory of the instrument. The successes of the instrument so far reported are encouraging.

Criticisms of the Inventory

While no criticisms of the Inventory of Choices have appeared as yet in the literature, one might expect some criticism on philosophical or psychological grounds. The authors of the instrument, however, do not argue that the dimensions of the Inventory are the most relevant in accounting for _all_ behavior, but feel that they are particularly relevant for an understanding of attitudes toward school subjects: the Inventory provides a useful model for schematizing behavior in terms which cut across motivational, learning, and perceptual categories.

In all studies except that of Maykovich (1966) an inherent assumption has been that Inventory scores remain stable over a reasonable period of time; Maykovich has shown that this assumption is not true for a long-term situation. The present investigation provides a check on the short-term stability of the attitude scores.

The content of a few items is apparently not familiar to some present-day high school students; for instance, Longfellow's poem "Evangeline" (item no. 20), and the nature of The Atlantic Monthly magazine (item no. 64). Furthermore, some items pose a problem to students of lower reading ability.

---

[8] Personal communication, January 28, 1960.

The answer sheet does not lend itself to rapid machine-processing. The method of indicating responses on a separate sheet, a procedure developed by the present writer, while facilitating keypunching of the responses, does not shorten processing time greatly: handscoring is complicated and tedious whether the original format or the present writer's answer sheet is used. While the present methods of indicating responses would suffice for studies using a relatively small number of subjects, a large-scale administration of the Inventory would be better recorded and processed on mark-sense cards or on answer sheets suitable for optical scanning.

On the technical side some criticisms are justified. The present form of the Inventory uses a six-choice continuum on which the response is located; this makes no provision for "undecided," "no opinion," or similar responses. Some students object to having to make a definite decision on a certain item and either decline to answer that item or else indicate more than one response. On the other hand, it may be argued that provision for "undecided" or "no opinion" types of responses encourages subjects to mark this type of response rather than to take time to consider the item carefully. Methods of weighting such neutral responses have been proposed, but have not been applied in scoring the Inventory.

A difficulty arises in applying the H-technique to the Guttman scales used in the Inventory. Non-cumulative response patterns, which are construed as errors of measurement, cause a loss of approximately 10% of subjects. While this loss is unavoidable, it is a drawback nonetheless. A Guttman scaling program[9] has been written for the IBM 704 computer; in this program no loss of subjects is incurred. The writer's data were processed by this program, and the results compared with those obtained by the H-technique. However, the use of such a program would not be practical when a small number of subjects is studied. There is also the danger that results reported by the computer seem to be more accurate than they really are, finer distinctions being implied than can be actually measured with such an instrument as the Inventory.

The Inventory of Choices booklet and answer sheet used in the present study are included as Appendix B.

---

[9] GUTS, written by Wm. Schutz, Department of Education, University of California, Berkeley; IBM SHARE Library No. 1337.

Summary

The cardinal aim of science education is the development of a scientifically literate citizenry. Six referents to scientific literacy have been isolated; these referents are illustrated by descriptions of the nature and aims of science such as that provided by Nagel. The historical development of the aims of science shows a diminishing importance attached to factual knowledge as an objective of science instruction. Despite the superficial differences found in various statements of the goals, there seems to be general agreement as to what science education should accomplish; the goals of chemistry instruction are not fundamentally different from those of science education in general.

The agreement in the literature notwithstanding, science teaching practices, especially testing, reveal that a gap exists between the stated objectives and classroom procedures. Teachers still appear to be testing to a substantial extent for facts, as are commercially-available tests; the practice is by no means restricted to science education.

Comparatively little work has been done on the components of chemistry achievement; all but a few recent tests combine various objectives into a single score rather than scoring the objectives separately.

Of several schemes proposed for classifying educational objectives, Bloom's Taxonomy of Educational Objectives (Cognitive Domain) has been researched the most thoroughly. This taxonomy is stated in behavioral terms and the objectives are assumed to pervade all curricula. The Taxonomy has been found appropriate for classifying the cognitive objectives of science teaching.

There seems to be sufficient agreement among judges in assigning test items to the main categories of the Taxonomy; empirical validation procedures generally have supported the cumulative and hierarchical structure and the distinctive nature of the four lower main categories, but not of the subdivisions within a category. On the other hand, the assumption of generality of process (behavior) over content (subject matter) has not been substantiated; the hierarchical and cumulative structure assumed of the Taxonomy gives rise to statistical problems which are as yet unresolved. Two attempts to identify patterns of achievement based on levels of the Taxonomy have not been successful.

The Inventory of Choices, a two-way classification of interests by Edwards and Wilson, has shown promise as a device for measuring the attitudinal orientation of high school students and as a predictor of academic achievement.

Both the Taxonomy and the Inventory of Choices are characterized by a clearly developed rationale. The Taxonomy seems manageable and sufficiently precise to serve as a classification scheme for the cognitive objectives of science education; the Inventory's scales have been constructed and refined with care. Both instruments have had their strengths and limitations revealed by research and thus are considered by the present writer to be suitable for playing a substantial role in the current investigation.

CHAPTER III

DESIGN OF THE STUDY

This study attempts to answer questions relating to patterns of academic achievement in secondary school chemistry. "Academic" is defined by Good (1959) as "pertaining to the realm of ideas or abstractions"; the present writer therefore interprets academic achievement as achievement of cognitive objectives, such as those outlined in the Taxonomy of Educational Objectives (Cognitive Domain). The most fruitful method of investigation would seem to be to study patterns of achievement by grouping items into subtests on the basis of common cognitive characteristics. The Taxonomy seems suitable for this purpose, for each major category could be represented by a subtest. Scores on the subtests then would constitute a student's achievement profile. Not only would students differ in the subtest scores they obtain, but it is conceivable that certain patterns might be identified with various groups of students. A suitable method of profile analysis would enable an investigator to determine whether a group of profiles could be considered sufficiently homogeneous to form a distinctive pattern; the method would also make possible the differentiation of patterns varying in shape, in level, or in both.

The Theoretical Framework of Pattern Analysis

The method of pattern analysis used in this study is that proposed by Haggard (1958, chap. VII). Haggard's method is an application of two-way analysis of variance in which subtests constitute one independent variable or main effect, and individuals constitute the other independent variable or main effect. In other words the scores are arranged in a c x k matrix with c subtest scores for k individuals. In the model employed a constant set of subtests will be used but the k individuals will vary since the number of individuals is specific to the group whose profiles are being studied. R_p (which is discussed later) is used as a measure of profile similarity.

Definition of Terms

Profile—the configuration of scores of an individual on a fixed set of variables comprising a set of subtests.

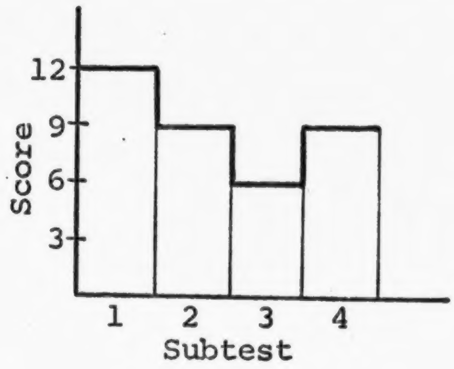


Fig. 2a--Profile of Individual A.

Pattern—a set of profiles.

Congruent Pattern—one in which all individuals in the group have the same score on each subtest.

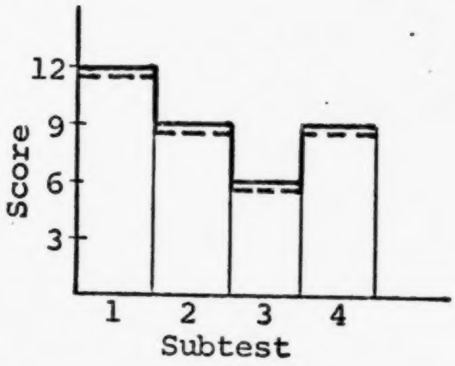


Fig. 2b--Congruent Pattern.
— Individual A
--- Individual B

$R_p = 1$

Parallel Pattern—one in which scores of individuals differ from on another by a constant, producing profiles which differ only in level. For this pattern $R_p = 1$ when differences in profile level (that is, among the k means) are partialled out.

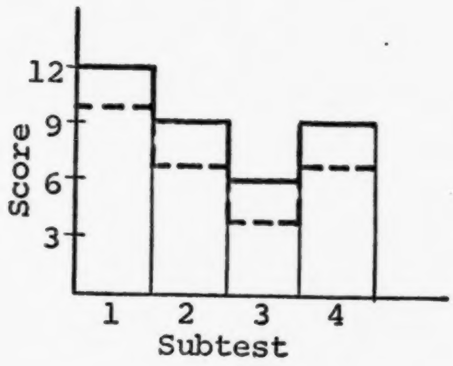


Fig. 2c--Parallel Pattern.
— Individual A
--- Individual B

$R_p = 1$

Mixed Pattern—one in which no identity or similarity occurs in the profiles. The overlap of profiles gives rise to a pattern resembling a random collection of profiles.

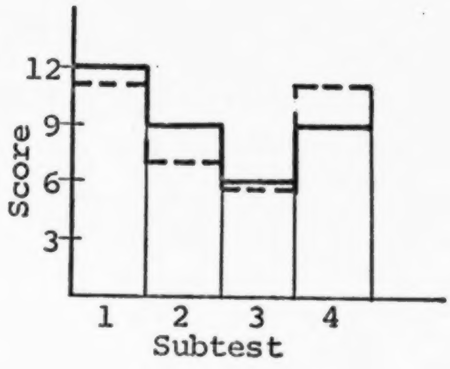


Fig. 2d--Mixed Pattern.
— Individual A
--- Individual B

$R_p = 0$

Incongruent Pattern—one in which the set of profiles is maximally dissimilar. For this pattern R_p reaches its minimum value of $\frac{-1}{k-1}$ where k is the number of individuals.

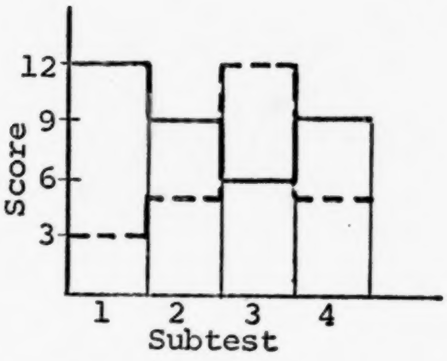


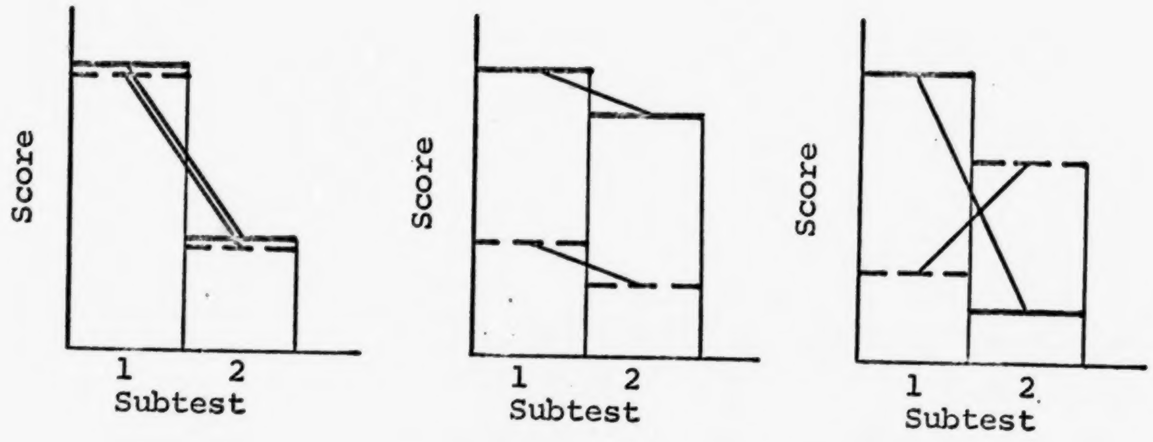
Fig. 2e--Incongruent Pattern.
—— Individual A
--- Individual B
$R_p = \frac{-1}{k-1}$

Fig. 2--Illustration of the Terms Profile and Pattern.

It should be noted that the above descriptions apply to ideal forms of patterns, and that the last two patterns are not considered patterns in the conventional sense. The concept of the incongruent pattern as distinct from the mixed pattern seems useful only when small numbers of individuals (say two to five) are under consideration. When k = 50, for example, the minimum value of R_p is approximately -.02 which is hardly different from zero.

The Function of Interaction in Profile Analysis

Consider the following profiles (in which the number of subtests has been reduced to two for simplicity):


(a) Congruent Pattern

(b) Parallel Pattern

(c) Either mixed or Incongruent Patterns

Subtest Differences:	Significant	Not Significant	Significant
Individual Differences:	Not Significant	Significant	Significant
Interaction:	Not Significant	Not Significant	Significant

Fig. 3--Interaction as an Indicator of Profile Dissimilarity

In Figures 3(a) and 3(b) one profile closely follows the trend of the other, as shown by the sloping lines. In Figure 3(c) in addition to significant differences in the two main effects, there is a significant combined interacting of the two main effects which is indicated by the crossing of the two trend lines. When this interaction is present no common trend in the two profiles is evident. Figure 3 illustrates the principle which is central to Haggard's method of pattern analysis: a congruent or parallel pattern is considered to exist when person-subtest interaction for a set of profiles is negligible or non-significant.

In terms of the F ratios obtained in the two-way analysis of variance a non-significant F for person-subtest interaction means a significantly congruent or parallel set of profiles; conversely a significant F for the interaction means that a pattern does not exist, in the conventional sense, among the profiles.

Two important points should be noted:

1. The above concept may be extended to any number of subtests.
2. Provided that certain conditions elaborated below are met, the technique can be used to compare patterns of different groups and to test the significance of their differences.

Prerequisite Treatment of the Data

Stabilizing of Scores

In the analysis of variance model used in profile analysis the estimate of the error variance (σ_e^2) is based on the variation among randomly sampled units within a cell, but since the number of scores per cell is one it is not possible to obtain an estimate of σ_e^2 from the data being analyzed.

Mental testing theory assumes that the error term for a given subtest has the same standard deviation (σ_e) over time for all individuals. If this assumption is true, σ_e is equivalent to the standard error of measurement of the subtest and may be estimated from the standard deviation of the subtest (σ_t) and its test-retest reliability coefficient (r_{tt}). The standard error of measurement, which is equal to $\sigma_t \sqrt{(1 - r_{tt})}$, provides an estimate of σ_e for that subtest; σ_e is assumed homogeneous over time for individuals.

When dealing with subtests simultaneously the subtest scores should be stabilized so that each subtest has the same variance over temporal repetitions. This stabilization is accomplished by dividing each score by its standard error of measurement; the resulting scores are termed stabilized scores (S scores). The effect of this procedure is to put the standard error of measurement equal to unity so that $\sigma_e = 1$ and $\sigma_e^2 = 1$ over repetitions for all subtests and individuals. This estimate of σ_e^2 is used as the residual or error mean square in the analysis of variance table. The effect of this procedure is to minimize the variance of scores over occasions and provide the most efficient estimator of overall effect of all the subtests taken together. It should be noted that since the estimate is derived from the test-retest reliability coefficient, that statistic must be obtained empirically or estimated from sources outside the data to be analyzed.

Additional Requirements for Comparison of Group Patterns

Haggard (1958, pp.118-122) points out that if his procedure is to be used for comparing groups rather than individuals the subtest scores must be:

1) normalized standardized scores based on the responses of a large population,
2) standardized on the same scale,
3) obtained from uncorrelated subtests.

Haggard admits that it is unsafe to assume that the majority of standardized tests meet these conditions. It seems especially unlikely that the third condition can be met except in the case where subtests have been deliberately constructed by orthogonal factor analytic methods to meet the third requirement.

A Further Requirement in the Present Study

It is expected that the criterion measure (chemistry achievement test) will be moderately correlated with scholastic aptitude and therefore it is proposed to examine the profiles formed from the residual scores rather than from scores which include scholastic aptitude as a significant component. This departure necessates a corrected value for the standard error of measurement of each subtest. This matter will be elaborated upon in Chapter V.

The Model, Formulas, and a Property of R_p

The two-way analysis of variance model used is a mixed model with C (subtests) fixed and K (individuals) random. Since the residual mean square is used as the denominator of the between-individuals and interaction F-ratios calculated in a two-way analysis of variance table, these F ratios become numerically equal to their respective mean squares when the residual mean square is put equal to unity by stabilizing the scores. The number of degrees of freedom for the residual mean square is equal to the number of subjects used in finding the test-retest reliability.

Using stabilized scores the two-way analysis of variance table reduces to the form shown in Table 2.

TABLE 2

ANALYSIS OF VARIANCE MODEL USED IN HAGGARD'S METHOD OF PATTERN ANALYSIS

Classification	df	MS	Expected Values of Mean Squares Under the Mixed Model
C(subtests)	(c - 1)	BCMS	$\sigma_e^2 + \sigma_{ck}^2 + k\sigma_c^2$
K(individuals)	(k - 1)	BKMS	$\sigma_e^2 + c\sigma_k^2$
C x K = I (interaction)	(c - 1)(k - 1)	IMS	$\sigma_e^2 + \sigma_{ck}^2$
R(residual)	(N for r_{tt})	(RMS)	$\sigma_e^2 = 1$

In Table 2

BCMS is the between-classes mean square

BKMS is the between-individuals mean square

IMS is the interaction mean square

RMS is the residual mean square

c is the number of subtests

and k is the number of individuals in the sample.

The appropriate F's and R_p may be calculated as follows:

$$F_{ck} = \frac{IMS}{1}$$

$$F_k = F_L = \frac{BKMS}{1}$$

$$R_p = 1 - \frac{IMS - 1}{IMS}$$

where F_{ck} is the test for significance of the c x k interaction,

$F_k = F_L$ is the test for significance of differences among the k means (that is, the differences in level among profiles)

and R_p is the measure of profile similarity.

Since, under the model assumed in this analysis, the expected value of IMS is $\sigma_e^2 + \sigma_{ck}^2$, the components entering the formula for R_p take on the form

$$1 - \frac{\sigma_e^2 + \sigma_{ck}^2 - 1}{\sigma_e^2 + \sigma_{ck}^2}$$

In the sample under consideration the residual variance has been equated to 1. With measures having an arbitrary metric, the numerical value of the variance is arbitrary and hence one may also put $\sigma_e^2 = 1$. The form of R_p then reduces to

$$R_p = 1 - \frac{\sigma_{ck}^2}{1 + \sigma_{ck}}$$

The ratio of interaction variance to the sum of error variance and interaction variance indicates the measure of profile dissimilarity. As σ_{ck}^2 increases, R_p approaches zero; when σ_{ck}^2 approaches zero, R_p approaches unity. R_p thus has a range of 0 to + 1, with the + 1 indicating complete profile similarity; the range and other properties of R_p therefore differ from the much more familiar Pearson product-moment correlation coefficient (Haggard, 1958, chap.III).

Since $R_p = 1 - \frac{\text{IMS} - 1}{\text{IMS}}$ and $\text{IMS} = F_{ck}$,

then $R_p = \frac{\text{IMS} - \text{IMS} + 1}{\text{IMS}} = \frac{1}{\text{IMS}}$;

therefore $R_p = \frac{1}{F_{ck}}$.

One property of R_p is especially noteworthy: for any given numbers of degrees of freedom the value of F at the .01 probability level is greater than its value at the .05 level. Since R_p is the reciprocal of F_{ck} the value of R_p at the .01 probability level is less than its value at the .05 level.

Another way of accommodating this unexpected property of R_p is to remember that in looking for the presence of congruent or parallel patterns, one is looking for absence of interaction and since some interaction will be present due to sampling errors, the .05 probability level for F_{ck} is, in this instance, a more stringent criterion than the .01 level for the interaction F. The smaller the value of F, the larger the value of R_p.

Procedure Used to Analyze Patterns

Consider subtest scores for one individual forming a profile, and also consider a number of profiles grouped according to some external criterion. Two questions now suggest themselves:

Question 1: Does each group have a pattern?

First each group must be considered separately. A null hypothesis is postulated that in the group a random arrangement of profiles exists, that is, that the profiles in the group have no similarity. The alternative hypothesis would then be that the profiles are not mixed but are congruent or parallel:

H_0: $R_p = 0$

H_1: $R_p \neq 0$ where R_p is the measure of profile similarity.

F_{ck} and R_p are now calculated.

Suppose that F_{ck} (the test for significance of the interaction of subtests and persons) is not significant: then R_p is significant. The null hypothesis H_0 would now be rejected and the conclusion drawn that the profiles are congruent or parallel to the extent indicated by R_p. To determine whether the pattern of this group is congruent or parallel a hypothesis is now postulated that no difference in level exists among the profiles in the group. The alternative to this null hypothesis is that a difference in levels does in fact exist:

$H_0: D_L = 0$

$H_1: D_L \neq 0$ Where D_L is a convenient symbol for the difference in profile or pattern levels.

F_L is now calculated.

Suppose that F_L (the significance test for differences in level) is significant. The null hypothesis H_0 would now be rejected and the conclusion drawn that the patterns which exist in the group are parallel. Should F_L be not significant, the null hypothesis would be accepted, and the conclusion drawn that the patterns were congruent. So far it has been established that for one group a pattern does or does not exist. The above procedure would now be repeated for the other groups.

Question 2: If patterns among two or more groups are observed, do the patterns differ significantly?

F_{ck} has been computed for each group and in each case has been found to be not significant. The profiles from, say, two patterns would be combined and a single analysis made to determine whether the patterns differ from each other.

Combining the two patterns, and repeating the procedures outlined above might yield a combined F_{ck} which is significant (that is, a combined R_p which is not significant). It would then be concluded that two different patterns do exist. If the combined interaction F_{ck} is not significant (and therefore the combined R_p is significant) then the two patterns are alike. In the latter case a combined F_L will indicate whether the patterns are congruent or parallel (that is, whether $D_L = 0$ or $D_L \neq 0$).

Where different patterns are shown to exist, the configuration of means of each of the subtest scores of the profiles in the pattern serves to identify each pattern.

Factors Assumed to Influence Chemistry Achievement

Scholastic aptitude is known to be moderately correlated with academic achievement whether measured by objective tests or teacher-assigned grades. It is with factors other than academic aptitude that this section is concerned.

Anderson (1950, 1949) found several factors related to student achievement in chemistry; most of these factors concerned the teacher rather than the student. Because of the differences in the educational systems of Ontario and Minnesota (where Anderson conducted his study) some of the factors have little relevance to the present study. However, teacher qualifications and experience were found by Anderson to be significantly related to student achievement in chemistry, and are studied in the present investigation. Class size, use of a laboratory manual in the chemistry class, and the student's educational plans were also found to be significantly related to the student's chemistry achievement.

The studies of Edwards and Wilson reviewed in the preceding chapter show that achievement is related to the student's attitudinal orientation. It is reasonable to suppose that a student's attitude toward school may also influence his achievement; furthermore achievement in chemistry may be related to the subjects which the student likes most or least.

In addition to the factors listed, certain characteristics of the pupil (such as his age, sex, occupational and educational aspirations) and his family background (such as number of siblings, the socioeconomic status as determined by the occupations of father and mother, the language spoken in the home, and the length of the family's residence in Ontario) conceivably may have some bearing on the student's achievement. To these factors we may add such factors as the textbook in use at the school, the length of the classroom period, the number of chemistry class periods per week, and whether or not the student is repeating the course in Grade 12 Chemistry.

Along with the teacher variables studied by Anderson one would expect that sex of the teacher, and the teacher's use of audio-visual aids, programmed instructional materials and other supplementary materials might influence student achievement in chemistry. One other important factor might be the personal qualities possessed by a given teacher; however it is not the intention of this study to analyze teacher personality variables.

It should be noted that personal, attitudinal and environmental factors studied in this investigation cannot be assigned unequivocally to one or other of these categories; when such factors are stated broadly, most of them could be argued to belong to at least two categories.

Delimitation

In the present investigation the sample studied consisted of pupils enrolled in Grade 12 Chemistry of the General Course, who were attending secondary schools in Ontario which had been selected at random.

Academic achievement in chemistry was measured by a specially constructed test, the Ontario Test of Achievement in Chemistry (OTAC) which measures achievement as subtest scores in four cognitive areas defined by the Taxonomy of Educational Objectives:

1.00 Knowledge
2.00 Comprehension
3.00 Application
4.00 Analysis

Outcomes in the affective and psychomotor domains of the

Taxonomy were considered beyond the scope of this study.

Personal, attitudinal and environmental factors studied in relation to academic achievement in chemistry were the following:

1) scholastic aptitude as measured by the Scholastic Aptitude Test, Ontario Edition (SATO);
2) relative achievement (overachievement, normal achievement, underachievement) expressed as the discrepancy between obtained and predicted OTAC scores;
3) the specialization of attitudes as measured by the scales of the Inventory of Choices of Edwards and Wilson;
4) sex of student;
5) educational plans and occupational aspiration of the student;
6) family data and home environment;
7) characteristics of the chemistry teacher;
8) characteristics of the school environment.[1]

Two other scholastic factors were studied in conjunction with the factors listed, namely final mark in chemistry (not including OTAC scores) and average Grade 12 final examination mark (not including final chemistry mark).

General Hypotheses

1. Total OTAC achievement scores show substantial corre-

---

[1] Detailed specification of each of these factors is found in Appendix H where complete lists of all variables used in the study are found.

lation (.40 - .60) with scholastic aptitude, the final mark in chemistry, and the average final examination mark. OTAC subtest scores show low correlation (.20 - .40) with these measures and with each other.

2. Patterns of achievement, measured as configurations of OTAC subtest scores with scholastic aptitude held constant, exist.
3. Patterns of achievement are related to the specialization of attitudes, educational aspirations and plans, family data, and other personal and environmental factors considered singly, when scholastic aptitude is held constant. Some interactions among the factors are present.

By using the Prudent, Theoretic, and Immediate poles to delimit those Inventory of Choices scales found relevant to achievement, and to serve as examples, it is possible to elaborate on section 3 above. When congruent or parallel patterns have emerged across the complete range of a scale such as the Prudent-Theoretic scale, it is reasonable to postulate that, because of their interests, Prudent students would be more proficient in the lower taxonomic skills in chemistry than would Theoretic students; one might postulate a Prudent pattern in which the mean score in Category 1.00 would be the largest, with means descending in magnitude as they increase in taxonomic level. A Theoretic pattern would be opposite in aspect, but higher in overall level, as indicated by a higher grand mean. Immediate students might be expected to follow the Prudent pattern but at a lower level than that of Prudent students. The differences between all pairs of means in a pattern would have to be tested for significance, since it is possible that small "steps" in a profile might be misleading.

While the Inventory of Choices scales have been used to illustrate the construction of hypothesized patterns, it is possible to postulate patterns on the basis of many of the other variables used in this study. Whenever patterns emerge across the whole range of a variable, such postulated patterns will be set forth.

Hypotheses Used in the Statistical Analysis

General hypotheses subsumed under section 1 above will be tested by inspection; the nature of the present study does not indicate a more rigorous testing of these hypotheses. With the exception of the hypothesis concerning interactions which are detected by a computer program described in Chapter V, the remaining general hypotheses are expressed symbolically in the hypotheses which follow. It should be noted that each statement of a hypothesis actually represents a family of hypotheses since many identifying variables are used in the present study.

For convenience a superscript in parentheses indicates the ordinal position of the hypothesis, while a subscript 0 indicates a null hypothesis and other subscripts indicate alternative hypotheses; for example, $H_P^{(5)}$ indicates an alternative to the fifth null hypothesis.

To test for the presence of a pattern in one group

$H_0^{(1)}$: $R_P = 0$

$H_1^{(1)}$: $R_P \neq 0$

$H_0^{(2)}$: $D_L = 0$

$H_1^{(2)}$: $D_L \neq 0$

To test whether two or more patterns differ in shape or level

$H_0^{(3)}$: $R_P = 0$

$H_1^{(3)}$: $R_P \neq 0$

$H_0^{(4)}$: $D_L = 0$

$H_1^{(4)}$: $D_L \neq 0$

To test whether emergent patterns have a characteristic shape and level

$H_0^{(5)}$: $M_1 = M_2 = M_3 = M_4$

$H_P^{(5)}$: $M_1 > M_2 > M_3 > M_4$

$H_I^{(5)}$: $M_1 > M_2 > M_3 > M_4$

$H_T^{(5)}$: $M_1 < M_2 < M_3 < M_4$

where P, I, and T represent Prudent, Immediate, or Theoretic on Inventory of Choices continuums.

$H_0^{(6)}$: $M_i = M_j$; $i \neq j$, $i = 1, 2, 3$
$j = i + 1, \ldots, 4$

$H_1^{(6)}$: $M_i \neq M_j$ where i and j designate the numbers of the subtests.

$H_0^{(7)}$: $\overline{M}_A = \overline{M}_B$ where $\overline{M}$ is the grand mean, and A and B designate two groups in the same continuum.

$H_{PT}^{(7)}$: $\overline{M}_T > \overline{M}_P$

$H_{TI}^{(7)}$: $\overline{M}_T > \overline{M}_I$

$H_{PI}^{(7)}$: $\overline{M}_P > \overline{M}_I$

CHAPTER IV

CONDUCT OF THE STUDY

The Criterion Instrument

The Ontario Test of Achievement in Chemistry (OTAC) is an end-of-course[1] test designed to measure achievement in the topics of the Ontario Grade 12 Chemistry syllabus. The present writer's twelve years' experience in teaching Grade 12 Chemistry, his experience with standardized tests, and the revision of three prototype tests were utilized in developing OTAC. The version of OTAC used in the present study forms Appendix F.

Experience with Published Tests

After careful consideration of all available published tests in secondary school chemistry, the author chose the Anderson Chemistry Test as the instrument having highest curricular validity with respect to the Ontario Grade 12 Course of Studies in Chemistry. The author conducted some pilot studies with this test in his school from 1959 to 1962. For three consecutive years Form Am of this test formed part of the final examination in chemistry, as well as being administered at other times (along with Form Bm). While this test provided much useful information in some retention and motivation studies, and proved particularly useful in grade placement of students entering the school from other provinces and countries, it was not judged suitable to be used alone for end-of-course evaluation or for the purposes of the present study. The following are reasons for this decision:

---

[1] To make possible the administration of the test in mid-May rather than in early June, the last few topics on the syllabus are omitted from the test; however, it is the present writer's experience that teachers generally cover more than 90% of the syllabus by mid-May.

1. Curricular validity was not sufficently high (19 of the 80 items deal with topics not on the Ontario Grade 12 syllabus).
2. The present writer classified the items of one form of the Anderson test according to the *Taxonomy*. A disproportionate number of items were concerned with factual material alone (39 of the 61 items having curricular validity were considered to be in Category 1.00 of the *Taxonomy*; this number represents 64% of these remaining items). In terms of the Ontario Grade 12 Course of Studies in Chemistry, five items were found to be in Category 2.00 and none in Category 4.00.
3. An item analysis made by the present writer showed point-biserial correlations of less than .20 for 18 of the 61 items having curricular validity.

Consequently a program to develop a more suitable instrument for use in Ontario was initiated. The development procedures which culminated in the present version of OTAC are outlined in Appendix D.

The Structure of OTAC

In building a test based on the *Taxonomy* the following points were kept in mind:

1. Test items must be assigned to *Taxonomy* categories only after carefully considering the educational experience of the pupils for whom the test is intended. (This educational experience is likely to be more uniform in Ontario, where a centrally prescribed course of study is followed, than in areas where courses of study vary considerably from community to community. Hence the use of the *Taxonomy* may be more appropriate in Ontario than in some neighboring states.)
2. The proportions of items in various categories cannot be prescribed rigidly. Which categories predominate depends on the program of studies followed, and, to a considerable degree, on the examiner's interpretation of the program as well as his experience in teaching it.

For these reasons the author's study does not test pupil proficiency in Categories 5.00 and 6.00 of the *Taxonomy*. In the author's opinion, these objectives represent levels of cognition forming a small portion of the total objectives of a course in introductory chemistry for secondary schools. In addition, items at these advanced levels require, or should allow for, some element of student originality; this requirement makes the usual objective-type item somewhat less appropriate. This view is shared by Dressel.[2]

Based on the amount of factual material introduced in chemistry at the Grade 12 level in Ontario, and the writer's experience in teaching the course in Ontario, the following distribution of items by categories seems most reasonable: Knowledge—24 items (40%), Comprehension—12 items (20%), Application—12 items (20%), and Analysis—12 items (20%). This distribution is close to those adopted by some investigators whose work is reported in Chapter II.

A problem in the construction of *Taxonomy*-type tests arises from the hierarchical and cumulative structure of the *Taxonomy*: cognitive categories higher than Knowledge cannot be tested independently of knowledge, and thus a student's score on Categories 2.00 and above depends on his ability to assimilate factual material and recall it on demand.

One method of avoiding the influence of factual knowledge on higher cognitive performance is to "hold content constant" by providing the student with the content in the form of a reading passage. This was the method employed in constructing Taxonomy-type tests by Kropp and Stoker (1966), Anderson (1964), and Herron (1966). In the present writer's opinion such a practice must place a premium on reading comprehension, and student differences in this area of competence may unduly affect their chemistry scores. The solution to this problem seems to be to administer a suitable reading comprehension test and adjust for this factor by analysis of covariance. None of the above investigators followed this practice.

---

[2] Personal communication dated November 11, 1962.

The present investigator did not attempt to hold content constant by incorporating reading passages in the test. It was felt that in the Ontario Grade 12 Chemistry course content was controlled, to a much greater extent than would be the case in most American school systems, by the rigid prescription of topics in the syllabus plus the fact that only two textbooks were approved for use in the publicly supported high schools which constitute the majority of high schools in Ontario. These features would guarantee a high degree of homogeneity of content, at least in what is taught throughout the province.

One of the dangers of not holding content constant is that the "pyramid" type of test may result, in which the higher categories of the Taxonomy are dependent on the success of the student on increasing numbers of items in lower categories. An example illustrates this point: a Category 4.00 item conceivably may not be successfully answered unless a student is able to answer two Category 3.00 items, which in turn depend upon successful completion of perhaps four Category 2.00 items; these in turn may depend on the successful completion of (or the possession of the knowledge equivalent to), say, eight or more Category 1.00 items. This pyramid effect may be reduced by choosing in the higher categories items which do not depend on the factual content of items at lower levels. This is the procedure adopted by the present writer in constructing OTAC: in constructing the process-by-content "test blueprint," care was taken to avoid as much as possible overlapping of content in the four categories. It must be admitted that the higher categories still require factual knowledge and that in scoring an incorrect response to an item in one of the higher levels, one does not know whether the item has been answered incorrectly because the student could not function at that level, or because some essential factual information could not be recalled. On the other hand, as teaching experience reveals, and results of reading comprehension tests confirm, presenting written factual material to a student does not guarantee that the student will notice all of it, assimilate it completely, or be able to recall all of it on demand.

It is worthy of note that Winter _et al_. (1965) did not use the reading passage design in constructing the _Taxonomy_-type tests used in their study.

It should be noted also that in constructing OTAC the present writer allocated items to various categories solely on the basis of his own experience in teaching chemistry. However, before the results were analyzed a panel of judges was used to finalize the allocation of items. This procedure resulted in a slight change in the proportion of items assigned to each level of the Taxonomy but this change in proportion did not affect appreciably the outcome of the analysis.

Selection of the Sample

In April of 1964, the Ontario Department of Education was approached for permission to canvass the secondary schools; consent was given the present writer to solicit 50 schools. Selection of the schools was made by tossing a die to choose a section of a table of random numbers (Arkin and Colton, 1950, Table 23). A location within the table was picked at random and a sequence of three-digit numbers was then recorded. The Department of Educational Research[3] made available to the writer its master list of secondary schools, each identified by a three-digit number. The three-digit numbers obtained from the random number table were then matched against the school numbers to select schools for the study. Inactive numbers were ignored.

---

[3] The Departments of Educational Research and Graduate Studies of the Ontario College of Education were reorganized as The Ontario Institute for Studies in Education on July 1, 1965.

Once the schools had been selected letters were sent to the principals soliciting their cooperation. Included in the first mailing was a letter from Dr. George E. Flower, Director of Graduate Studies of the Ontario College of Education, supporting the request. A two-page explanation of the study was also enclosed along with a reply sheet. Senior administrative officials of the larger school systems were sent copies of the first mailing. Administrators, department heads, and teachers with whom the writer was acquainted were approached personally where possible, or by telephone or letter in an attempt to encourage cooperation.

Of the 50 schools approached, 31 (62%) agreed to participate in the study, 16 (32%) declined, and only 3 (6%) failed to reply.

A complete set of all materials mailed to schools is included as Appendix D.

Data Collection Procedures

On May 6, 1964, participating schools were shipped the following: an acknowledgment, test booklets, mark-sense answer cards, questionnaire booklets and answer sheets, administration instructions, test and questionnaire report forms, and return instructions. Administration of OTAC took place between May 11 - 26, while the Inventory of Choices and personal information questionnaire were completed between May 8 - 26. A specimen of the personal information questionnaire comprises Appendix C.

On return of the testing and other materials a questionnaire was sent out to be completed by each chemistry teacher in the school. This questionnaire provided much of the needed information concerning teacher and school characteristics.

OTAC answer cards were processed on the Department of Educational Research IBM 1401 computer.

On May 30 a list containing the OTAC scores and percentile ranks of students was posted to the schools; included in this mailing were two tables for converting percentile ranks to school marks, and a note explaining their use.

On June 19 another mailing was sent to the schools. This mailing consisted of nominal rolls prepared on the computer for the entering of final examination results and related information together with letters of thanks from the writer and from Dr. Flower.

Concurrent with the main data collection described above a test-retest administration was conducted in the writer's school; this school was not one of those selected at random for the study. The first administration of OTAC in the writer's school took place on May 21. The retest was given as part of the student's final examination on June 12. In anticipation of the progress of the study the Inventory of Choices had been administered to all Grade 12

Chemistry students in the writer's school on October 22, 1963. The Inventory was administered again to these students on May 27, 1964 in order to gain some estimate of its stability.

Preparation of the Data for Analysis

In the autumn of 1964 all data from the schools were checked and the few omissions and transcribing errors corrected. Much of the information collected had to be transferred to punched cards and considerable time was spent in carrying out this operation.

The responses to the Inventory of Choices were keypunched directly from the answer sheets and checked by the writer. In May of 1965 the writer visited Dr. T. Bentley Edwards at Berkeley, California and discussed with him several research problems relating to the Inventory of Choices. At Dr. Edward's suggestion, the responses to the Inventory were scored using the Guttman scaling program at the University of California's computer installation at Berkeley.

In June of 1965 the scores of the 1963-64 edition of the Scholastic Aptitude Test (Ontario Edition)[4] were transferred to the writer's decks from the master file of the Department of Educational Research. It was discovered that no 1963-64 SATO scores were available for about 14% of the students in the present study, and it was considered worthwhile to retrieve as many of these missing scores as possible. The computer output of the 1962-63 edition of SATO was searched but only one-sixth of the students whose scores were missing could be identified with certainty. Letters were then sent to the participating schools asking them to provide scores from any administration of SATO that might be in their records. By this time scores from the 1964-65 edition of SATO were available for some students who had not written the previous year's edition, and these were returned by the schools together with the results from the 1962-63 and 1961-62 editions for other students. The end result of this retrieval was that missing SATO scores were reduced from approximately 14% to about 4%; it was necessary, however, to equate the scores from four different editions of SATO if all these retrieved scores were to be useful. The details of this equating process are given in Appendix J.

---

[4] Administered in November, 1963.

In September of 1965 the Inventory of Choices was rescored, this time according to Stouffer's H-technique, by the present writer. An IBM 519 document-originating machine and IBM 84 high-speed sorter were used to unscramble the items, score them, form compound items, and gangpunch these and the scale scores into IBM cards.

The writer was fortunate to receive from the University of Chicago Statistics Laboratory a test-scoring and item analysis program written in FORTRAN II by Arieh Lewy. This program was rewritten in FORTRAN IV and made compatible with the University of Toronto's IBM 7094 computer by the present writer, and then used to score the subtests and perform the item analysis on the OTAC responses. For this program the OTAC responses had to be converted from the mark-sense split-field punch pattern to the more conventional format used by the Lewy program.

Personal information obtained from the student questionnaire (Appendix C), school marks and related information were transcribed by clerks to mark-sense cards; these cards were spot-checked and processed by the writer. Some of the information in this deck required all 12 punch positions in some columns. To make the data in this deck more amenable to computer analysis, a program was written to convert zone punches to values of "11" and "12"; two consistency checks were incorporated into the program so that absurd combinations of data were detected (for example, a student indicating that he planned to leave school without completing Grade 12 and then enter a university). A few such unlikely combinations were found and traced to clerical miscoding.

In order to avoid spurious correlation between average marks and chemistry marks, and between OTAC scores and final chemistry marks which might contain OTAC as a component, a program was written by the present writer to remove the respective component in each case. The formula used to remove the chemistry mark from the average mark was

$$Q = \frac{YZ - X}{Z - 1}$$

where Q is the average final mark after removing the final chemistry mark,

Y is the average mark including the chemistry mark,

Z is the number of curriculum subjects used to compute the average, and

X is the final chemistry mark of the student.

In this formula all marks are expressed in percent.

Since each school was free to determine to what extent the OTAC score should enter into the computation of the student's final chemistry mark, different formulas were used for various schools to remove the OTAC component, if any, from the final chemistry mark. The following formula was used for two schools which simply used the example given by the writer in the Chemistry Mark List (Appendix D).

$$G = \frac{10}{9} B$$

where G is the final chemistry mark, in per cent, after removal of the OTAC mark, and

B is the chemistry mark less the OTAC mark. In this example B has a maximum value of 90, since the OTAC mark formed 10% of the final chemistry mark.

Of the 30[5] schools in the study, 15 did not use OTAC scores as part of the final Chemistry mark, 2 used OTAC scores only to assist borderline students at the annual promotion meetings, and 13 used OTAC scores converted to a school mark to form part of the final Chemistry mark. Where used, the OTAC scores comprised from 4% to 50% of the final mark, with a median value of 10%.

Information concerning each teacher and each school was coded and punched into a teacher deck. This information was then gangpunched into the students' cards.

All data from the various decks were finally transferred to a master deck which contained two IBM cards per student. A slightly modified version of the master deck was prepared on the University of Toronto IBM 7094 computer to meet the requirements of the AID program which was run at the University of Michigan.

Appendix H contains lists of all variables available and the coding systems used for those variables requiring it.

---

[5] After the schools had been selected it was discovered that two of the schools participating in the study had amalgamated, thus reducing the number of participating schools from 31 to 30.

CHAPTER V

ANALYSIS OF THE DATA AND DISCUSSION OF FINDINGS

Characteristics of the Sample

The sample consisted of 2,339 Grade 12 Chemistry students who wrote the Ontario Test of Achievement in Chemistry. Scholastic Aptitude Test (Ontario Edition) scores were available for 2,248 of these students. A comparison of the sample statistics with those of the Ontario Grade 12 General Course population is given in Table 3.

TABLE 3

SATO SAMPLE AND POPULATION STATISTICS

Test	Group	N	Mean	SD
SATO TV	Chemistry sample	2,248	26.18	8.33
SATO TV	General Course	44,029	26.50	8.29
SATO MATH	Chemistry sample	2,248	17.05	5.69
SATO MATH	General Course	44,016	17.18	5.69

It is apparent that the sample statistics are quite close to the General Course population parameters, but it should be kept in mind that the sample cannot be considered as having been drawn randomly from that population since only about 64% of the population enrolled in Grade 12

Chemistry. In contrast, the students writing OTAC represent 81% of the Grade 12 General Course enrollment in the schools participating in the study.

The percentages of boys and girls writing OTAC were 52.6 and 47.3 respectively. Of those writing the criterion test, 9.7% were taking the Grade 12 Chemistry course for the second time; in this group 67.1% of the repeaters (6.5% of the sample) were boys.

The average chemistry student in the sample was 17.8 years of age. Most students had a favorable attitude toward school, and intended to proceed to university after completing Grade 13. The largest group of students came from families which had resided in Ontario for over 100 years; almost as large a group came from families which had resided in Ontario from 10 to 24 years.

The mean length of chemistry period in the schools in the sample was 37.5 minutes; the mean number of periods per week devoted to chemistry was 5.4. The average chemistry class contained 30.4 students.

Of the 48 chemistry teachers in the sample, 39 (81.4%) were males. The mean Grade 12 Chemistry teaching experience for this group of teachers was 9.5 years and their mean Grade 13 Chemistry teaching experience was 4.7 years.

A summary of personal, attitudinal, and environmental data collected comprises Appendix H: for continuous variables, means and standard deviations are calculated; for some variables, medians and frequency distributions are provided.

Characteristics of OTAC

Conventional Test Statistics

OTAC was scored with no correction made for guessing. The criterion test statistics are given in detail in Appendix G. Table 4 presents those statistics of major interest.

Examination of Table 4 shows that OTAC on the whole was difficult for the group; the subtest containing Category 4.00 items was especially difficult. The overall variability of OTAC is quite acceptable (Ebel, 1965, p.302). Median point-biserial correlations of items to their own subtests are in every case higher than the corresponding correlations to the test as a whole, suggesting that the subtests have some measure of uniqueness. (If items correlated no better to their own subtest than they did to the test as a whole, one could doubt that the arrangement of items into those subtests had any valid basis.)

The reliability of OTAC is low compared to the reliabilities reported for published chemistry tests. Of the factors tending to depress test reliability (Mursell, 1949, p.46-50; Ebel, 1965, pp.336-338, 343-344), the following are most likely to have affected OTAC:

1. Too many difficult items;
2. A wide range of item difficulties which causes the

TABLE 4

Main Item Statistics for OTAC

Category	Number of Items	Score Range	Mean	S.D.	Difficulty		Point-Biserial Correlation				Reliability
							To Subtest		To Total Test		
					Range	Mean	Range	Median	Range	Median	KR-20
Total	60	3-56	25.15	8.13	.11-.79	.42	.20-.49	.36	.14-.43	.29	.819
1.00	23	0-23	9.71	3.50	.11-.79	.42	.20-.43	.34	.16-.37	.29	.640
2.00	11	0-11	5.25	2.32	.23-.59	.48	.38-.49	.44	.27-.41	.34	.570
3.00	14	0-14	6.19	2.60	.24-.74	.44	.26-.49	.40	.17-.43	.32	.590
4.00	12	0-10	4.00	1.87	.13-.49	.33	.26-.41	.34	.14-.36	.20	.315

Kuder-Richardson Formula 20 (KR-20) reliability to be underestimated;

3. The presence of items of low discriminating power;
4. Heterogeneous items: this factor must account substantially for the lower reliability observed, since the test was deliberately constructed of subtests purporting to measure different cognitive abilities;
5. Items not scaled in order of difficulty;
6. The comparatively small number of items (60) as compared to published chemistry tests.

The figures in Table 4 support the first three reasons. The fifth reason is supported by examination of the detailed item analysis in Appendix G.

Administration irregularities can depress the reliability of a test. Only two schools reported conditions which distracted students who were taking the test: one school reported a room temperature of 80°F and another school had its students writing the test in a cafeteria where noise from the kitchen bothered the testees. The mood and attitude of students taking the test, and the varying motivation of the students, are known to affect test reliability; unfortunately it is impossible to assess the effect of such factors in the administration of OTAC. It should be noted that all objective tests are susceptible to these conditions however, and there is no reason to believe that OTAC was administered under conditions very much different from those encountered during the administration of other objective tests such as SATO.

According to a rather arbitrary but commonly used criterion (Swineford, 1966, p.5), a test is regarded as essentially unspeeded if at least 80% of the group reach the last item and if virtually everyone reaches 75% of the items. Of the group writing OTAC, 88% reached the last item and only two students (less than 0.1%) did not reach at least three-quarters of the items in OTAC. These figures indicate that in all likelihood speed was not an important factor in the test.

The low reliabilities of the subtests may be due to a number of causes, one of which is the small number of items. It is useful to calculate what the reliabilities would be if each subtest consisted of 60 items whose kind and quality were similar to that of the actual items used (see Table 5). For this purpose, the Spearman-Brown "prophecy formula" (Cronbach, 1960, p.131) is useful:

$$r_n = \frac{nr}{1 + (n - 1)r}$$

where r is the original reliability

and r_n is the reliability of the test n times as long as the original test

TABLE 5

PROJECTED RELIABILITIES OF OTAC SUBTESTS

Subtest	n	Observed KR-20	KR-20 projected on basis of 60 items in each subtest
1	60/23	.640	.823
2	60/11	.570	.878
3	60/14	.590	.860
4	60/12	.315	.697

The small number of items is definitely one cause of low subtest reliability. Other causes contributing to low subtest reliability seem to be the same as those that affected the test as a whole. The heterogeneity of items within each subtest is probably made more pronounced by the attempt to avoid as much as possible stereotyped item formats in assembling the test.

The much lower reliability of subtest 4.00 is difficult to explain, but may be attributed in part to the unfamiliar format of many of the items in this category. Ontario secondary school students on the whole have not been exposed to formal objective tests in science to the same extent as many of their counterparts in other places; the relative unsophistication of such students may result in such items causing consternation during the test. Any testing irrelevancy such as a distraction of this sort would tend to depress the reliability of the subtest.

It is worth noting that of all subtests, subtest 4.00

had the highest percentage (11.2%) of omitted items.

A test-retest administration of OTAC was given to the students of the present writer's school in order to obtain the necessary data for stabilizing the test scores; the major statistics of this administration are presented in Table 6.

It should be noted that the retest could not be considered to be given under the same conditions as the test. The retest formed part of the students' final chemistry examination when motivation was probably at the highest level one could expect. On the other hand, it is the writer's experience that some students do not prepare for a test such as OTAC as seriously as they would for a final examination. One might reasonably expect, then, to find a reflection of this change to more uniform motivation in the item statistics of the test and retest. Examination of Table 6 shows that compared to the test, the retest has higher means as well as higher standard deviations and reliability coefficients (except in Category 3.00).[1] The shift to higher means and reliability coefficients may be due also to practice effect (although the students did not know that the same test would be given twice), but the one significant change in variability is an increase rather than a decrease as expected.

---

[1] All gains in means were significant at the .01 level; changes in variance were not significant except for the gain in Category 1.00, which was significant at the .05 level. Appropriate t tests for correlated observations are described in Popham (1967, p.152) and Ferguson (1966, pp.183-184).

TABLE 6

Test-Retest Statistics for OTAC

Category	Number of Items	Pretest 27 May 64 N = 185			Posttest 12 June 64 N = 187			Test-Retest Reliability
		Mean	S.D.	Reliability (KR-20)	Mean	S.D.	Reliability (KR-20)	N = 182
Total	60	27.88	7.49	.789	32.13	8.09	.820	.821
1.00	23	10.42	2.99	.543	11.87	3.33	.607	.692
2.00	11	6.00	2.31	.572	6.65	2.36	.627	.665
3.00	14	7.00	2.78	.656	8.32	2.63	.622	.715
4.00	12	4.45	1.88	.254	5.30	1.98	.342	.315

In all cases except that of subtest 4.00, the test-retest reliability exceeds the corresponding KR-20 reliability obtained in the main OTAC administration and in the test-retest administration.

The point-biserial correlations of nine OTAC items to the total test fell below the arbitrary limit of .20. An examination of these items and their statistics revealed no strong reason for rejecting any of them. In deciding to retain these items in the analyses, the writer was guided by the suggestions of Cronbach (1960, pp.366-367).

In general, although the test proved to be difficult for the group to which it was administered, it compares favorably with published tests. The KR-20 reliability of the test as a whole, although not as high as many similar reliabilities reported for published tests, seems quite acceptable, particularly when one considers that the test attempts to measure distinct cognitive abilities plus a wide range of factual matter loosely brought together under the heading "chemistry achievement" (Ebel, 1965, p.336). Davis (1964, p.23-24) points out that, for large groups, reliabilities lower than .50 may still be useful; however, his reminder to take into account the varying reliabilities of tests applies equally to subtests. The low reliability of subtest 4.00 especially must be kept in mind, since its standard error of measurement (a function of the subtest's reliability) will be large and the sensitivity of that subtest therefore will be decreased.

Characteristics Peculiar to Taxonomy-Type Tests

Subtest difficulty and Taxonomy category

In a test constructed on the principles of the Taxonomy of Educational Objectives, one would expect that as the level of cognitive ability becomes more complex the mean score of the corresponding subtest decreases. This inverse relationship between category level and mean score generally has been observed in studies reported by Kropp and Stoker (1966, pp.82-84).

Examination of subtest difficulties in Table 4 shows no such tendency. The reason is that in building OTAC an attempt was made to equalize the difficulty of the subtests. Items vary in difficulty for reasons other than their membership in a certain *Taxonomy* category. A good example of this variation is found in the Category 1.00 subtest in OTAC, where item difficulties range from .11 to .79. In an item pool there could exist a wide range of item difficulties in each category, and it should be possible to select, either inadvertently (particularly where items are tied to subject matter and course coverage) or deliberately, subtests of high or low average difficulty, and thus prove or disprove the relationship of increasing difficulty with category. One may circumvent the possibility of biasing results in this manner by using items whose difficulty is either concealed or not known. On the other hand, it seems advantageous to construct Taxonomy subtests equal in average difficulty so that student preferences for various cognitive functions may reveal themselves; in other words one may attempt to "keep subtest difficulty constant" in investigating the cognitive preference of students. This was the approach taken in assembling OTAC. As shown in Table 4, the attempt was not altogether successful. The lack of constant difficulty from subtest to subtest can be attributed to the fact that only a relatively small item pool was available from which to choose items to build the subtests and that the group on which the items had been pretested was not representative of the group which wrote OTAC.

It should be noted that in the preliminary editions of OTAC, the higher the Taxonomy category, the more difficult the items were on the average. Appendix E illustrates the foregoing relationships of mean item difficulty and Taxonomy category membership.

Simplex Structure

If the subtests in a test form a hierarchical structure, one would expect that adjacent subtests would be more highly correlated than would subtests whose positions are remote in the hierarchy. The subtest intercorrelation matrix should assume an aspect similar to that shown in

Table 7. Any matrix in which diagonal entries are relatively high and in which entries drop in value as they are situated farther away from the diagonal is said to possess simplical structure or to form a simplex.

TABLE 7

EXAMPLE OF A PERFECT SIMPLEX

Subtest	1	2	3	4
1	-	.74	.55	.42
2		-	.74	.56
3			-	.76
4				-

While real data seldom fit a mathematical model, it is possible to obtain a measure of the goodness of fit of the data to the model; the statistic q^2 suggested by Kaiser (1962, pp.155-162) is such a measure. Table 8 gives the subtest intercorrelation matrix obtained for OTAC when the items were assigned to _Taxonomy_ categories by the panel of judges.

TABLE 8

OTAC SUBTEST INTERCORRELATION MATRIX

Taxonomy Category	1.00	2.00	3.00	4.00
1.00	-	.538	.548	.408
2.00		-	.577	.387
3.00			-	.401
4.00				-

The matrix shown in Table 8 does not form a perfect simplex, but sufficiently approaches one to lend support to the hierarchical structure amongst the subtests. Using Kaiser's method, a q^2 of .90 is obtained for the simplex-like matrix in Table 8.[2] This value of q^2 compares very favorably with values for a chemistry test reported by Kropp and Stoker (1966, p.87).

Agreement of Judges

Research mentioned in Chapter II indicates that reasonable agreement can be expected amongst qualified judges when they assign items to Taxonomy categories. In the present study three teachers of chemistry from three different publicly supported high school systems served as judges; all had several years' experience in teaching Grade 12 and Grade 13 Chemistry. The present writer had originally assigned OTAC items to Taxonomy categories while assembling the test in 1964. A year later the items were again classified by the writer, but without reference to the first classification. A few months later the judges met with the writer, were provided with a synopsis similar to Table 1, and, after an explanation of the Taxonomy, were asked to classify the OTAC items. Much discussion took place but the judges classified items without knowing which classification was given to any item by the other judges or by the present writer. In order to avoid influencing panel members, the present writer was careful not to have his own classification available while the panel of judges sat; his function was restricted to clarifying issues related to the Taxonomy as a classification scheme. The panel submitted their classifications after a session lasting about six hours.

---

[2] The same operations performed on the intercorrelation matrix obtained when items assigned to Taxonomy categories by the present writer alone yield a q^2 of .91.

In assigning the OTAC items to Taxonomy categories for this study, the classifications submitted by the three judges were compared with the two classifications made a year apart by the present writer. The five sets of item classifications were pooled and the modal classification of each item determined; items were then assigned to Taxonomy categories according to these modal classifications. Table 9 shows the extent of agreement as indicated by modal classification of OTAC items.

The results reported in Table 9 surpass those reported by Stoker and Kropp (1964, pp.39-42) even when the present writer's classifications are omitted from the pool. When only the three judges' classifications are pooled, the results equal those reported by McGuire (1963b). The high degree of agreement amongst judges supports the argument that the Taxonomy categories are meaningful to item classifiers and that the OTAC subtests have some degree of judgmental validity.

TABLE 9

EXTENT OF INTER-JUDGE AGREEMENT IN ALLOCATING OTAC ITEMS TO TAXONOMY CATEGORIES

Number of Judges Agreeing on Placement of Items	Extent of Agreement or Disagreement	Number of Items Concerned
5	Unanimous	30
4	1 category difference[a]	12
4	2 categories difference[b]	5
4	3 categories difference[c]	1
3	1 category difference[a]	5
3	2 categories difference[b]	6
2 + 2[d]	2 categories difference[b]	1
	Total Number of Items.....	60

[a] Difference of opinion at most one category removed from majority opinion.

[b] Difference of opinion at most two categories removed from majority opinion.

[c] Difference of opinion at most three categories removed from majority opinion.

[d] In the one case where a bimodal response occurred between Categories 1.00 and 2.00, the higher category was selected.

Pyramid Effect

The danger of pyramid-like relationships occurring in Taxonomy-type tests has already been referred to in Chapter IV (pp.118-119). It was suspected that the unusual item statistics of the Category 4.00 subtest were the result of pyramid effects. To determine whether such pyramid effects existed in OTAC, the items were pasted on filing cards and then compared with one another. The investigator then attempted to place related items into pyramid structure. There were no groups of four or more related items spread across four different Taxonomy categories. One group of four items was spread across three categories, but the item in the highest category was the easiest item of the group—the opposite to the expected result. Two groups of three items each were spread across three categories, but the easiest item in both cases was in the middle category. Three groups of three items were found spread over two categories each, but two of these three groups fit the pyramid model only partially. Of nine item pairs detected, the higher category item in five pairs was the more difficult, as would be expected. In three pairs the higher category item was the less difficult of the pair, and in the remaining pair both items were of equal difficulty.

One may conclude that the pyramid effect in OTAC is negligible, and is probably the result, to a substantial extent, of the attempt to equalize item difficulty across Taxonomy subtests. In any case the item difficulties observed in the groups of related items suspected of pyramid effects do not seem to explain the unusual statistics of the Category 4.00 subtest.

The foregoing results show that OTAC possesses the characteristics of Taxonomy-type tests to a satisfactory degree. The inverse relationship between mean subtest score and Taxonomy level of subtest was deliberately modified in assembling the final edition of the test. Simplex structure and sufficient inter-judge agreement are evident; pyramid structure is virtually absent. The conventional test statistics show OTAC to be suitable for research purposes.

The present writer thus considers OTAC a satisfactory criterion test for the purposes of this investigation.

AID Analysis

In the present study a large number of variables assumed to be related to chemistry achievement are investigated; the process of identifying those variables and their interactions which are related to chemistry achievement is necessarily a complicated one requiring numerous repetitive trial-and-error procedures. The Automatic Interaction Detector (AID) program (Sonquist and Morgan, 1964) carries out the process by empirically selecting those combinations of independent variables which account for the variance of a dependent variable.

The AID program is written in MAD (Michigan Algorithmic Decoder), an Algol-based programming language used extensively at the University of Michigan.

The AID Program

The AID program is designed to select optimal combinations of explanatory variables. One dependent variable and up to 36 independent (explanatory) variables may be handled in one run.

The objective of the AID program is to explain the variance of the dependent variable. The program accomplishes this objective by using a series of binary splits to subdivide the sample into a series of mutually exclusive subgroups which maximize one's ability to predict values of the dependent variable; that is, each group's mean may be used to predict the scores of students falling in that group. The analysis uses a non-symmetrical branching technique based on the processes of variance analysis.

No assumptions of linearity, additivity, normality, or absence of interactions are required as in conventional multiple-regression techniques. Categorical data, rankings, and continuous variables may at the same time be used as predictors subject to certain restrictions.

Program Restrictions

Independent variables may be monotonic (ordered) or free (non-ordered), but must fall in the range $0 \leqq V \leqq 63$. Negative values of independent variables are not permitted. Free independent variables should be limited to six or fewer classes.

The Algorithm

The essential steps are as follows:

1. Calculate the total sum of squares (TSS_t) around the grand mean of the dependent variable.
2. Select for splitting that group i which has the largest total sum of squares of the dependent variable around

the group mean (TSS_i) and which satisfies parameters P1 and MSIZE mentioned below.

3. Split group i into two non-overlapping subgroups to provide the largest reduction in the unexplained sum of squares; that is, maximize the between-group sum of squares (BSS_i). This procedure is performed on group i over all possible binary splits on all predictors, with the proviso that $\frac{BSS_i}{TSS_i} \geqq$ parameter P2 for a split to occur.
4. Repeat step 3 with the next most promising group (the one with the next largest TSS_j) selected by step 2, and continue until the process is terminated in accordance with the parameters which follow.

Input Parameters

The values inserted here are those which were used in the actual runs, and selected on the advice of the originators of the program.

1. P1, Split Eligibility Criterion = 0.015; that is, 1.5% of TSS_t must be in a group if that group is to become a candidate for splitting.
2. P2, Split Reducibility Criterion = 0.005, that is, the best split on any group must reduce the unexplained sum of squares by at least 0.5% of TSS_t or that group will not be split and will not become a candidate for splitting even though it may satisfy parameter P1.

3. MAXGP = 50. This is the maximum number of final groups into which the input data may be split, regardless of parameters P1 and P2.
4. MSIZE = 25. This is the minimum number of observations (students) a group must have if that group is to become a candidate for splitting.

Of the parameters listed, P2 is the most crucial.

Thus a group will not be split further if: (a) less than 1.5% of TSS_t is present in that group, or (b) the unexplained sum of squares is not reduced by at least 0.5% of TSS_t by the best split, or (c) less than 25 students are present in the group.

The process terminates when all groups cannot be split further, or when more than 50 unsplit final groups appear.

Types of Groups

The series of binary splits produces a number of final groups which fall into three categories:

1. Small group—one which contains too few observations to warrant further splitting;
2. Explained group—one above the minimum size, but having too little variation to warrant further splitting;
3. Unexplained group—one which is sufficiently large and spread out, but for which no variable in the analysis was useful in reducing the unexplained variation within that group.

Other Features of the Program

Monotonic predictors have their order maintained during the partition scan. Free variables have their classes rearranged during the partition scan and are sorted in descending sequence according to the mean value of the dependent variable for each class.

Variables used in an early split are eligible for use in later splits; the program does not discard a variable after using it once. Thus, predictors can substitute for one another, at any stage, in explaining variation in the dependent variable.

Importance and Significance

The authors of the AID program base the splitting procedure on statistical importance. A split is important when it reduces the unexplained variation by a large amount, whereas a split is significant if it cannot be quite reasonably attributed to chance. The program proceeds by making splits primarily on explanatory power, although significance of each split is not neglected.

Chance splits can be minimized in this program (a) by leaving monotonic variables in their natural rank-ordering during the partitioning process,[3] and (b) by leaving free variables unconstrained and having five or six classes at most.

---

[3] An exception to this rule is made when the possibility of a U-shaped or inverted U-shaped relationship is suspected between the dependent variable and an independent variable. In this case adjoining classes should be combined to form up to five classes and the independent variable left unconstrained.

Interaction

The graphic representation of the branching process carried out by the program is called the tree pattern. Non-symmetry in the tree pattern (that is, the extent to which different variables are used in the splits on the various trunks of the tree) implies interaction. If a variable used on one of the trunks shows no actual or potential utility in reducing predictive error in another trunk, then an interaction effect between that variable and those used in the preceding splits is indicated.

Advantages of the AID Program

A splitting predictor may not function with equal effectiveness over a number of groups which are the result of splitting by another predictor. The region of the tree where the predictor was most effective is immediately apparent.

Variables that "almost made it" (that is, were almost as useful as the splitting variable at any stage) are easily detected since the potential usefulness of each predictor at every stage is indicated by the ratio $\left(\frac{BSS}{TSS}\right)_i$ printed in the output. This feature guards against an interpretation based only on those variables which were actually used in the split. The same feature reveals variables which had little explanatory power at any stage in the splitting process.

Method of Investigating Variables

Five AID runs were made on the IBM 7090 computer at the University of Michigan. The first run was intended to be exploratory and contained only 18 independent variables plus the OTAC Total score as the dependent variable. The 18 independent variables used were those thought most likely to be related to OTAC chemistry achievement. After the results had been examined it was decided to carry on this part of the investigation by (a) discarding those variables which had shown themselves to be of little use and substituting for them independent variables not yet investigated, and (b) repeating the process using the first set of independent variables but substituting as the dependent variable *Taxonomy* Category 4.00 subtest scores in one case and chemistry final examination marks in the other.

AID Output

From the numerical output of the AID program it is possible to construct a graphical display of the relationships of the dependent variable to those independent variables which proved useful in explaining the former's variance; such a display is called an "AID tree" and consists of "trunks," "branches," and "twigs."

The output also provides values of the statistic BSS_i (the between-group sum of squares) for each predictor over each group created during the partition process. Those variables which were used in the splits, and those which were almost effective enough to be used for splitting, can be identified readily when values of BSS_i are presented in tabular form.

A useful aid in interpreting the results of an AID analysis is a table summarizing the contributions, across all groups, of those variables which were found to be important in explaining the variance of the dependent variable.

In the present study the results of the AID analysis are presented in all three forms mentioned.

Results

The abbreviations for the independent variables which were used in the AID analysis are listed in Table 10. These abbreviations are used to save space in the following tables.

Run No. 1

As shown in Figure 4, a number of mutually exclusive subgroups are formed by the action of relatively few explanatory variables: only five out of the 18 independent variables used in this run account for the explainable variance in OTAC total scores. The following interactions are indicated: Educational Plans 1 x SATO Total Verbal, Prudent-Theoretic x SATO Mathematics, SATO Mathematics x SATO Total Verbal, Theoretic-Immediate x SATO Mathematics x

TABLE 10

ABBREVIATIONS FOR INDEPENDENT VARIABLES USED IN AID ANALYSIS

Abbreviation	Variable
SATO TV	Scholastic Aptitude Test (Ontario)—Total Verbal Score
SATO MATH	Scholastic Aptitude Test (Ontario)—Mathematics Score
P-T 4	Inventory of Choices, Prudent-Theoretic 4-point scale
P-I 4	Inventory of Choices, Prudent-Immediate 4-point scale
P-A 4	Inventory of Choices, Prudent-Aesthetic 4-point scale
T-I 4	Inventory of Choices, Theoretic-Immediate 4-point scale
T-A 4	Inventory of Choices, Theoretic-Aesthetic 4-point scale
A-I 4	Inventory of Choices, Aesthetic-Immediate 4-point scale
SEX	Sex of student
OCCF	Occupation of father
OCCM	Occupation of mother
OCCS	Occupational aspiration of student
REPEATING	Repeating Grade 12 Chemistry
ATTITUDE	Attitude toward school
LANGUAGE	Language spoken in the home
ED PLANS 1	Immediate educational plans
ED PLANS 2	Future educational plans
SCHOOL TYPE	Publicly supported, Roman Catholic, Independent private
TEXT	Textbook used in chemistry class
MANUAL	Laboratory manual used in chemistry class
P-T 9	Inventory of Choices, Prudent-Theoretic 9-point scale
P-I 9	Inventory of Choices, Prudent-Immediate 9-point scale
P-A 9	Inventory of Choices, Prudent-Aesthetic 9-point scale
T-I 9	Inventory of Choices, Theoretic-Immediate 9-point scale
T-A 9	Inventory of Choices, Theoretic-Aesthetic 9-point scale
A-I 9	Inventory of Choices, Aesthetic-Immediate 9-point scale
NOB	Number of older brothers in family
NYB	Number of younger brothers in family
NOS	Number of older sisters in family
NYS	Number of younger sisters in family

TABLE 10—CONTINUED

Abbreviation	Variable
NORPOS	Normal ordinal position of child in family
INPOS	Reverse ordinal position of child in family
NCHILN	Number of children in family
RESIDENCE	Length of residence in Ontario
BEST SUBJ	Subject liked best by student
WORST SUBJ	Subject liked least by student
AID'S	Use of audiovisual aids and autoinstructional devices
AGE	Age of student
P-T 12	Inventory of Choices, Prudent-Theoretic 12-point scale
P-I 12	Inventory of Choices, Prudent-Immediate 12-point scale
P-A 12	Inventory of Choices, Prudent-Aesthetic 12-point scale
T-I 12	Inventory of Choices, Theoretic-Immediate 12-point scale
T-A 12	Inventory of Choices, Theoretic-Aesthetic 12-point scale
A-I 12	Inventory of Choices, Aesthetic-Immediate 12-point scale
TCHR RESP	Responsibility of chemistry teacher
TCHR SEX	Sex of chemistry teacher
TCHR QUAL	Qualification of chemistry teacher
12 CHEM EXP	Number of years chemistry teacher has taught Grade 12 chemistry
13 CHEM EXP	Number of years chemistry teacher has taught Grade 13 chemistry
PD LENGTH	Length of chemistry period
PREPS	Number of preparations per week for chemistry teacher
12 CLASSES	Number of Grade 12 chemistry classes per week
13 CLASSES	Number of Grade 13 chemistry classes per week
TCHG PDS	Total number of teaching periods per week
TCHG TIME	Teaching time per week (all subjects)
TOT PUPILS	Total number of pupils per week
CHEM PD/WK	Number of periods per week allotted to Grade 12 chemistry
TCHR	Teacher identification
CLASS SIZE	Size of students' chemistry class

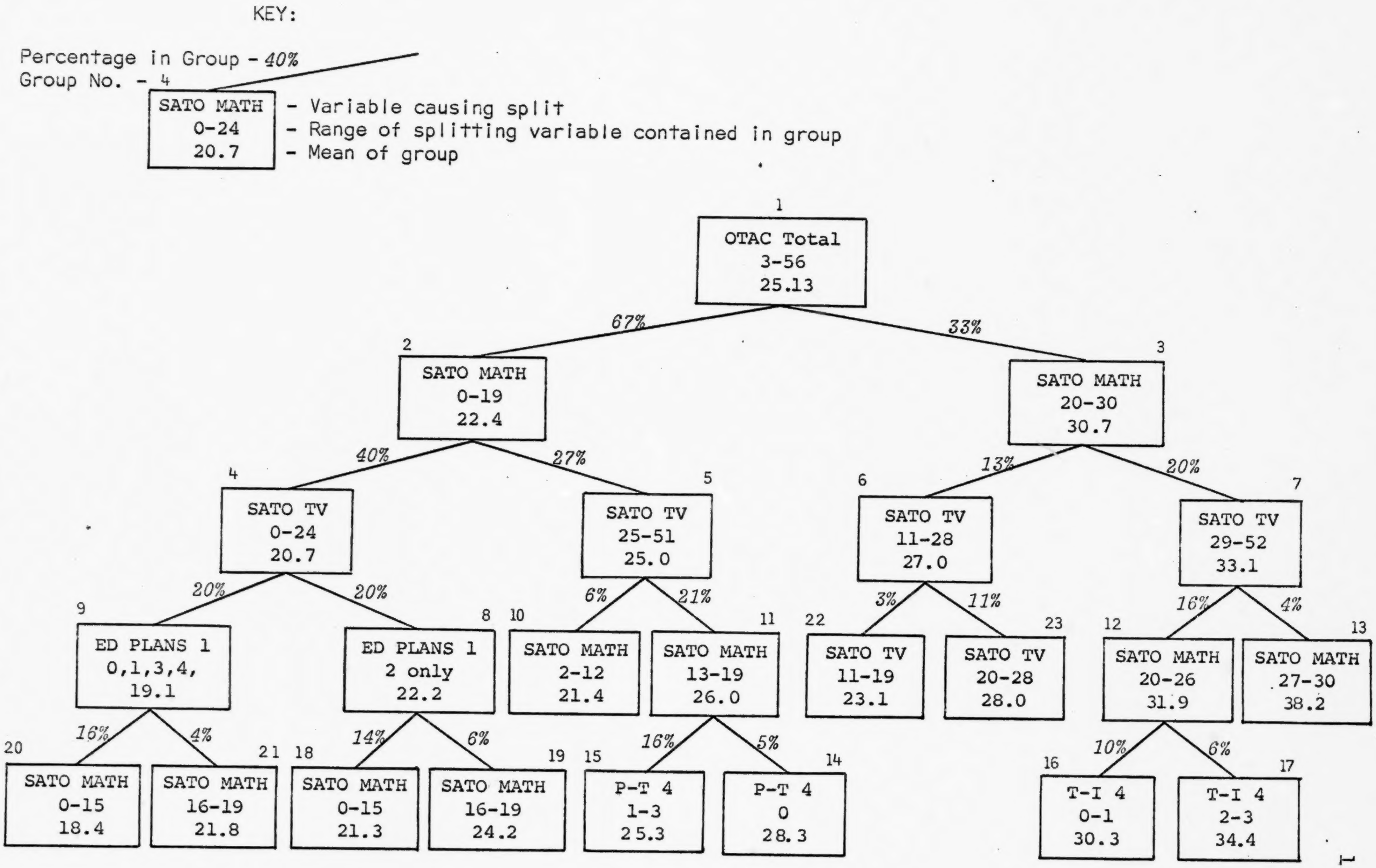

KEY:
Percentage in Group - 40%
Group No. - 4
SATO MATH
0-24
20.7
- Variable causing split
- Range of splitting variable contained in group
- Mean of group
1
OTAC Total
3-56
25.13
67%
33%
2
SATO MATH
0-19
22.4
3
SATO MATH
20-30
30.7
40%
27%
13%
20%
4
SATO TV
0-24
20.7
5
SATO TV
25-51
25.0
6
SATO TV
11-28
27.0
7
SATO TV
29-52
33.1
20%
20%
6%
21%
3%
11%
16%
4%
9
ED PLANS 1
0,1,3,4,
19.1
8
ED PLANS 1
2 only
22.2
10
SATO MATH
2-12
21.4
11
SATO MATH
13-19
26.0
22
SATO TV
11-19
23.1
23
SATO TV
20-28
28.0
12
SATO MATH
20-26
31.9
13
SATO MATH
27-30
38.2
16%
4%
14%
6%
16%
5%
10%
6%
20
SATO MATH
0-15
18.4
21
SATO MATH
16-19
21.8
18
SATO MATH
0-15
21.3
19
SATO MATH
16-19
24.2
15
P-T 4
1-3
25.3
14
P-T 4
0
28.3
16
T-I 4
0-1
30.3
17
T-I 4
2-3
34.4

FIG. 1 - AID Tree for OTAC Total Scores - Run No. 1

TABLE 11

AID ANALYSIS FOR RUN NO. 1 -
OTAC TOTAL SCORE
(Between-Group Sum of Squares for Each Predictor at Each Stage) × 10^2

	Variable	Group Number																						
		1	2	3	4	5	7	11	12	8	9	15*	6	22†	18*	20*	16*	23*	17*	14*	19*	10*	21*	13*
2	SATO TV	291	(72)	(68)	9	11	12	7	2	3	2	5	⑫		4	3	2	2	1	10	4	5	2	7
3	SATO MATH	(354)	72	50	22	㉑	㉘	6	7	⑧	⑨	5	5		4	3	3	5	3	1	0	2	0	2
4	P-T 4	31	8	12	3	10	17	⑧	1	1	3	0	4		2	3	3	3	3	1	0	1	2	3
5	P-I 4	114	2	7	0	0	0	0	1	0	1	0	2		1	0	1	2	2	0	1	0	1	0
6	P-A 4	11	4	2	3	1	3	1	4	1	1	0	1		0	1	2	1	0	2	0	0	0	0
7	T-I 4	42	6	21	5	2	21	2	⑯	2	4	3	4		2	3	2	3	1	1	1	0	1	3
8	T-A 4	36	4	15	2	1	13	0	10	1	1	1	2		2	1	5	1	1	0	0	2	0	1
9	A-I 4	11	2	6	0	0	3	0	2	0	0	1	0		1	0	0	0	2	1	1	1	1	1
10	SEX	68	8	6	9	5	6	0	5	5	5	1	4		3	3	1	3	3	0	0	3	0	0
11	OCCF	5	-	1	0	1	0	0	1	2	2	0	1		1	1	1	1	1	1	1	2	1	1
12	OCCM	6	5	1	1	1	0	2	1	0	2	1	1		2	2	1	2	1	0	1	2	0	1
13	OCCS	55	6	12	4	5	9	5	8	2	3	3	3		2	2	5	2	3	3	0	1	1	1
14	REPEATING	1	2	5	5	1	2	0	2	3	7 ◄	1	1		2	2	1	1	2	1	1	0	5	0
15	ATTITUDE	7	4	4	0	4	3	4	2	0	1	4	1		0	1	1	0	2	0	1	0	1	2
16	LANGUAGE	0	1	2	2	0	2	0	3	2	2	0	1		2	1	1	0	1	0	0	0	1	2
17	ED PLANS 1	147	47	27	㉓	6	8	4	9	-	1	3	8		-	1	6	3	3	1	-	1	1	0
18	SCHOOL TYPE	4	0	7	1	0	7	0	8	1	1	1	2		1	1	5	1	2	2	0	1	0	2
19	TEXT	0	1	0	0	0	0	1	1	0	0	0	0		0	1	1	0	0	1	0	0	0	0

* Unexplained Final Group

† Explained Final Group

○ Split made on this variable

◄ Points to next best splitting variable

SATO Total Verbal.

As indicated in Table 11, the variable "Repeating" "almost made it," that is, would be almost as good a splitter as SATO Mathematics for group 9. The total variance explained by each important variable in the run is shown in Table 12.

TABLE 12

AID ANALYSIS FOR RUN NO. 1—
CONTRIBUTION OF IMPORTANT VARIABLES
TO OTAC TOTAL SCORE VARIANCE

Variable	Percent of Variance Explained
SATO MATH	27.55
SATO TV	10.66
ED PLANS 1.	1.50
T-I 4	1.02
P-T 4	0.54
Total	41.27

Thus 41.27% of the variance of OTAC total scores is explained by five variables. None of the remaining variables used in this run reduced the unexplained sum of squares by as much as 0.5%.

Run No. 2

Variables which were of practically no use in explaining the variance of total OTAC scores were replaced with untried variables. The four-point Stouffer scales of the Inventory of Choices were replaced with the nine-point Guttman scales obtained from the computer runs made at the

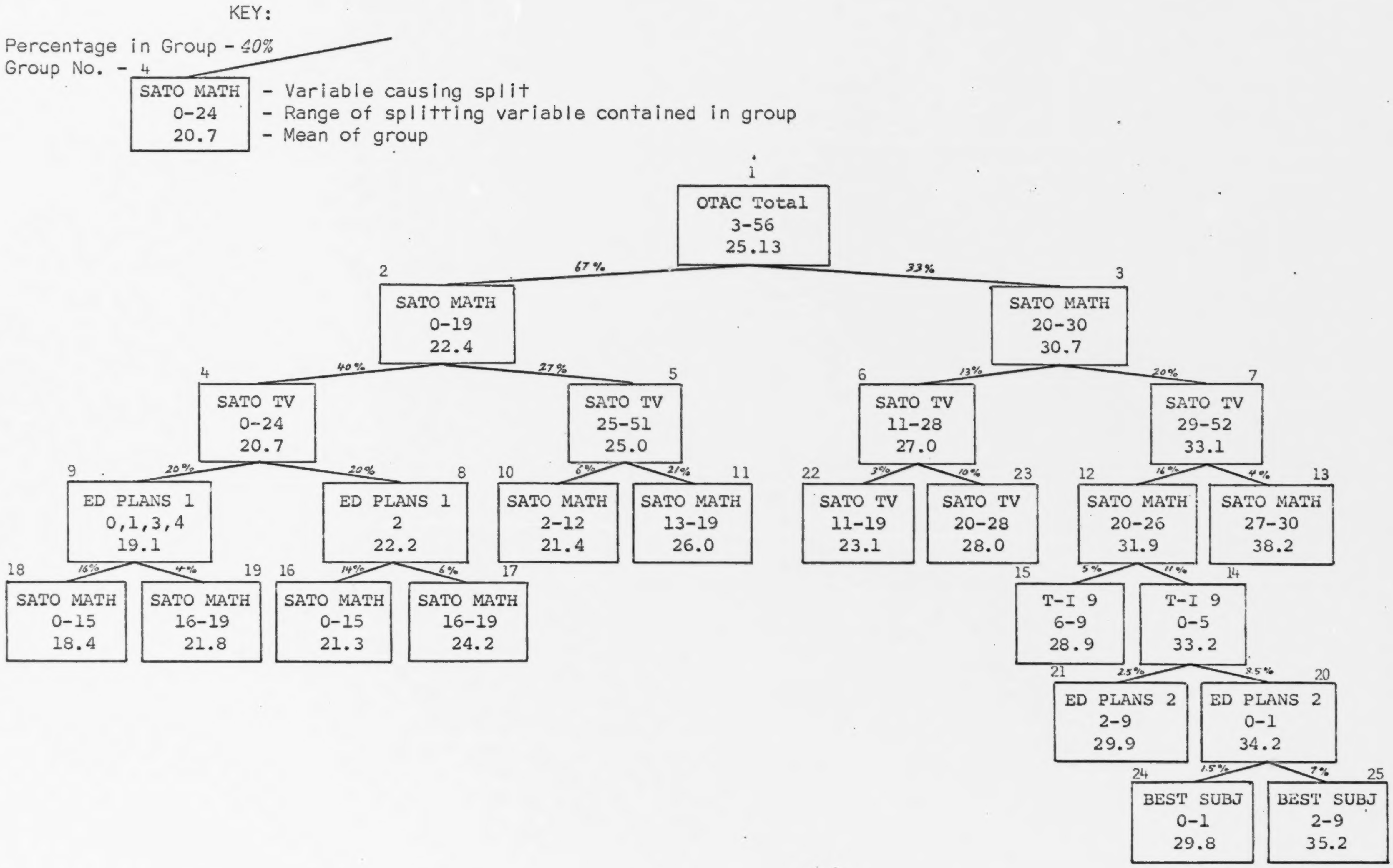

KEY:
Percentage in Group - 40%
Group No. - 4
SATO MATH 0-24 20.7
- Variable causing split
- Range of splitting variable contained in group
- Mean of group
1 OTAC Total 3-56 25.13
67%
33%
2 SATO MATH 0-19 22.4
3 SATO MATH 20-30 30.7
40%
27%
13%
20%
4 SATO TV 0-24 20.7
5 SATO TV 25-51 25.0
6 SATO TV 11-28 27.0
7 SATO TV 29-52 33.1
20%
20%
6%
21%
3%
10%
16%
4%
9 ED PLANS 1 0,1,3,4 19.1
8 ED PLANS 1 2 22.2
10 SATO MATH 2-12 21.4
11 SATO MATH 13-19 26.0
22 SATO TV 11-19 23.1
23 SATO TV 20-28 28.0
12 SATO MATH 20-26 31.9
13 SATO MATH 27-30 38.2
16%
4%
14%
6%
5%
11%
18 SATO MATH 0-15 18.4
19 SATO MATH 16-19 21.8
16 SATO MATH 0-15 21.3
17 SATO MATH 16-19 24.2
15 T-I 9 6-9 28.9
14 T-I 9 0-5 33.2
2.5%
3.5%
21 ED PLANS 2 2-9 29.9
20 ED PLANS 2 0-1 34.2
1.5%
7%
24 BEST SUBJ 0-1 29.8
25 BEST SUBJ 2-9 35.2

FIG. 5 - AID Tree for OTAC Total Scores - Run No. 2

TABLE 13

AID ANALYSIS FOR RUN NO. 2 -
OTAC TOTAL SCORE
(Between-Group Sum of Squares for Each Predictor at Each Stage) × 10^2

Variable		Group Number																									
		1	2	3	4	5	7	11*	12	8	9	14	21†	6	22†	16*	18*	20	24†	23*	25*	15*	17*	10*	19*	13*
2	SATO VERBAL	291	72	68	10	11	12	7	2	3	2	2		12		4	3	3		2	3	2	4	5	2	7
3	SATO MATH	354	72	50	22	21	28	6	7	8	9	4		5		4	3	5		5	2	2	0	2	3	2
4	P-T 9	49	5	9	5	9	13	7	9	6	1	7		4		4	0	6		3	3	1	1	1	1	2
5	P-I 9	9	3	5	2	1	2	0	3	1	1	2		2		1	1	2		1	4	2	0	1	3	1
6	P-A 9	18	8	8	4	0	4	1	3	1	4	4		2		1	4	5		1	1	1	1	0	1	2
7	T-I 9	36	7	17	4	4	22	4	15	2	2	1		3		3	3	1		3	4	2	1	1	1	3
8	T-A 9	65	18	15	6	6	11	5	7	2	6	3		1		2	3	1		1	1	2	2	1	3	3
9	A-I 9	12	8	3	2	2	3	2	2	1	2	2		2		1	2	2		1	1	1	1	1	1	3
10	SEX	68	8	6	9	5	6	1	5	5	5	5		4		3	3	7◄		3	2	0	0	3	0	0
11	NOB	9	3	2	1	2	1	1	2	0	1	1		1		0	1	0		1	0	1	0	0	0	0
12	NYB	5	2	6	1	1	0	1	1	1	1	3		1		1	1	1		1	1	1	1	1	1	3
13	NOS	4	1	1	0	2	1	0	1	0	1	2		0		0	1	0		0	0	1	2	1	1	0
14	NYS	49	5	9	5	9	13	7	9	6	1	7		4		4	0	6		3	3	1	1	1	1	2
15	NORPOS	8	3	2	1	1	1	1	1	1	1	0		1		0	1	1		1	1	1	1	0	1	1
16	INPOS	7	3	3	1	1	2	1	1	1	1	1		2		1	2	2		1	2	1	1	2	2	3
17	NCHILN	17	6	6	4	1	1	1	1	1	2	1		2		1	2	3		1	2	1	1	1	1	2
18	RESIDENCE	5	3	2	1	2	2	3	2	3	1	2		1		2	1	3		0	1	1	0	1	4	2
19	BEST SUBJ	38	4	8	3	10	13	6	9	3	1	8		3		3	1	8		2	2	0	1	0	1	2
20	WORST SUBJ	15	5	2	7	4	6	3	3	5	5	2		3		4	2	3		3	1	6	0	0	1	2
21	ED PLANS 1	147	47	27	23	6	8	4	9	-	1	5		8		-	1	4		3	3	3	-	1	1	0
22	ED PLANS 2	201	52	34	16	14	13	8	10	8	2	9		4		5	2	3		2	-	1	2	3	1	1
23	MANUAL	3	4	7	1	7	4	6	3	1	1	3		1		0	2	3		2	2	2	1	0	1	1
24	AID'S	27	14	7	6	6	2	6	2	1	5	3		2		2	3	4		2	3	0	1	0	1	1
25	AGE	25	4	15	1	3	6	4	4	0	2	3		2		0	2	3		1	1	1	1	3	1	1

* Unexplained Final Group
† Explained Final Group

○ Split made on this variable
◄ Points to next best splitting variable

University of California at Berkeley. SATO Total Verbal scores, SATO Mathematics scores, and Educational Plans 1 categories were retained in this run.

As shown in Figure 5, again a number of mutually exclusive subgroups are formed by the action of relatively few explanatory variables: only six out of the 24 independent variables used in this run account for the explainable variance in OTAC total scores. The following interactions are indicated: Educational Plans 1 x SATO Total Verbal, SATO Mathematics x SATO Total Verbal, Theoretic-Immediate 9 x SATO Mathematics, Educational Plans 2 x Theoretic-Immediate 9, Best Subject x Educational Plans 2.

As indicated in Table 13, the variable SEX would be almost as good a splitter as Best Subject for group 20. The total variance explained by each important variable in the run is given in Table 14.

TABLE 14

AID ANALYSIS FOR RUN NO. 2—
CONTRIBUTION OF IMPORTANT VARIABLES
TO OTAC TOTAL SCORE VARIANCE

Variable	Percent of Variance Explained
SATO MATH	27.55
SATO TV	10.01
ED PLANS 1.	1.50
T-I 9	0.95
ED PLANS 2.	0.57
BEST SUBJ	0.55
Total	41.13

Thus 41.13% of the variance of OTAC total scores is explained by six variables. None of the remaining variables used in this run reduced the unexplained sum of squares by as much as 0.5%.

It is worth noting (Table 14) that, in terms of total variance explained, Theoretic-Immediate 9, Educational Plans 2, and Best Subject together are almost as effective as Educational Plans 1 and Theoretic-Immediate 4 (Table 12) in accounting for the percentage of variance explained.

Run No. 3

The procedure of Runs No. 1 and 2 was repeated, with the twelve-point Guttman scales replacing the nine-point Guttman scales of the Inventory of Choices. Variables which were of practically no use in explaining the variance of total OTAC scores were replaced with untried variables.[4]

As shown in Figure 6, once again a number of mutually exclusive subgroups are formed by the action of relatively few explanatory variables: only six out of the 26 independent variables used in this run account for the explainable variance in OTAC total scores. The following interactions are indicated: Teacher x SATO Total Verbal, SATO Total Verbal x SATO Mathematics, SATO Mathematics x Teacher, Prudent-Theoretic x SATO Total

---

[4] Educational Plans 1 was inadvertently omitted from this run, because of a coding error in the control cards; however, the importance of this variable has been established in the preceding runs.

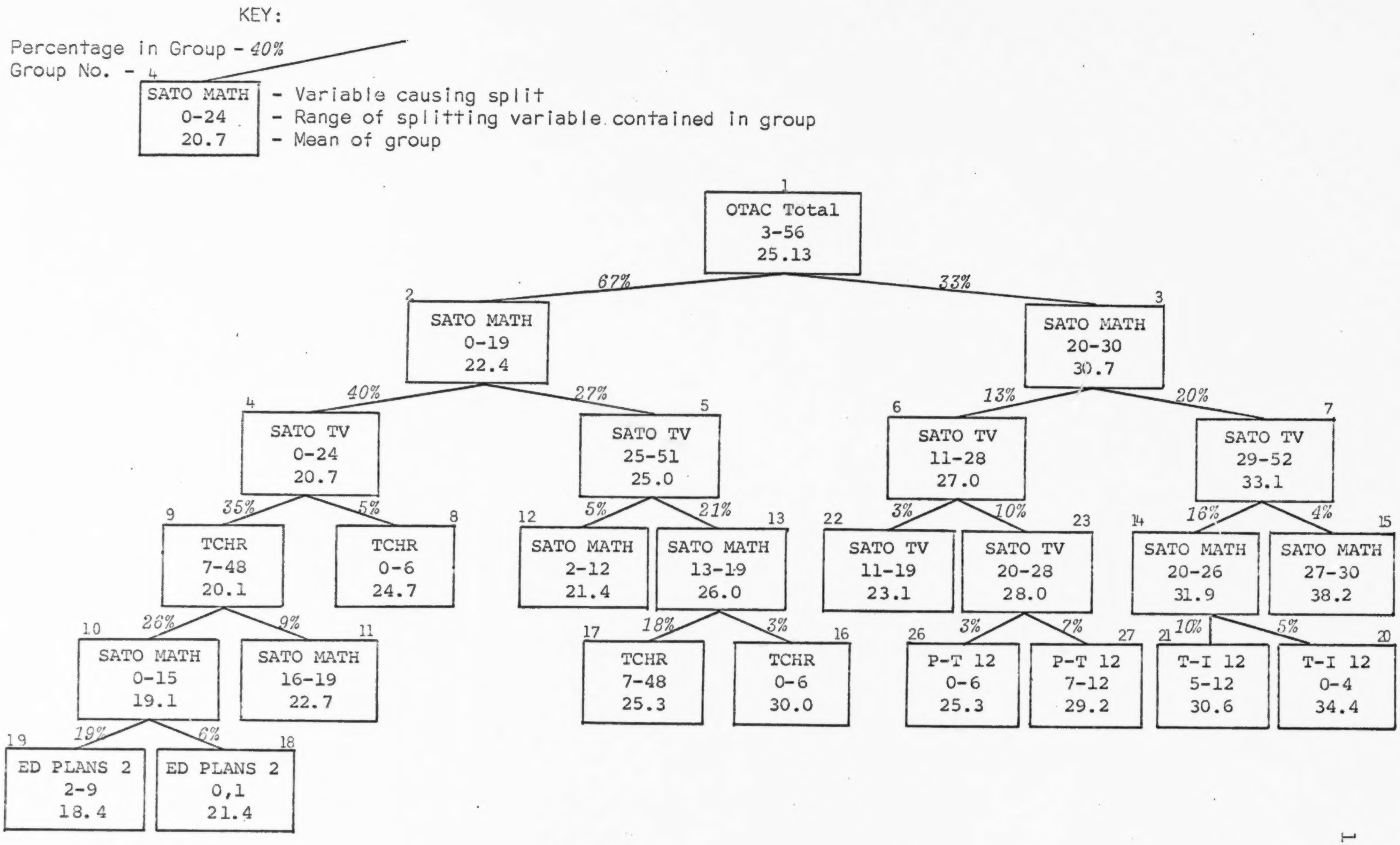

KEY:
Percentage in Group - 40%
Group No. - 4
SATO MATH
0-24
20.7
- Variable causing split
- Range of splitting variable contained in group
- Mean of group
1
OTAC Total
3-56
25.13
67%
33%
2
SATO MATH
0-19
22.4
3
SATO MATH
20-30
30.7
40%
27%
13%
20%
4
SATO TV
0-24
20.7
5
SATO TV
25-51
25.0
6
SATO TV
11-28
27.0
7
SATO TV
29-52
33.1
35%
5%
5%
21%
3%
10%
16%
4%
9
TCHR
7-48
20.1
8
TCHR
0-6
24.7
12
SATO MATH
2-12
21.4
13
SATO MATH
13-19
26.0
22
SATO TV
11-19
23.1
23
SATO TV
20-28
28.0
14
SATO MATH
20-26
31.9
15
SATO MATH
27-30
38.2
26%
9%
18%
3%
3%
7%
10%
5%
10
SATO MATH
0-15
19.1
11
SATO MATH
16-19
22.7
17
TCHR
7-48
25.3
16
TCHR
0-6
30.0
26
P-T 12
0-6
25.3
27
P-T 12
7-12
29.2
21
T-I 12
5-12
30.6
20
T-I 12
0-4
34.4
19%
6%
19
ED PLANS 2
2-9
18.4
18
ED PLANS 2
0,1
21.4

FIG. 6 - AID Tree for OTAC Total Scores - Run No. 3

TABLE 15

AID ANALYSIS FOR RUN NO. 3 -
OTAC TOTAL SCORE
(Between-Group Sum of Squares for Each Predictor at Each Stage) × 10^2

Variable		Group Number																								
		1	2	3	4	9	5	7	13	10	14	17*	6	19*	21*	22†	23	11*	20*	18*	27*	12*	8*	15*	26*	16*
2	SATO VERBAL	291	72	68	10	7	11	12	7	8	2	8	12	2	2		2	3	2	5	2	5	3	7	1	5
3	SATO MATH	354	72	50	22	20	21	28	6	8	7	7	5	2	6		5	0	3	5	4	2	1	2	1	1
4	P-T 12	70	15	13	10	9	11	16	7	5	10	7	8	4	5		8	1	4	6	0	1	2	2	1	0
5	P-I 12	15	7	8	4	4	1	4	2	2	3	1	1	1	2		3	3	2	1	1	1	1	2	1	1
6	P-A 12	4	0	2	1	1	2	1	1	1	1	1	4	0	2		4	1	2	2	2	0	2	0	1	0
7	T-I 12	64	17	18	12	12	6	15	6	9	12	4	5	7	4		5	3	2	1	5	1	1	2	1	2
8	T-A 12	51	16	7	8	6	4	7	3	3	4	3	2	1	4		2	1	2	1	1	2	1	2	2	1
9	A-I 12	10	4	6	1	1	1	1	2	1	3	1	2	1	2		2	1	2	2	1	1	1	1	2	1
10	SEX	68	8	6	9	8	5	6	1	4	5	0	4	0	1		3	0	5	5	2	3	0	0	0	1
11	ED PLANS 1	3	0	1	0	0	0	1	0	0	1	0	0	0	1		0	0	0	1	1	0	0	0	0	0
12	ED PLANS 2	201	53	35	16	17	14	13	8	11	10	5	4	1	4		2	3	5	2	1	3	1	1	0	2
13	TCHR RESP	27	9	12	9	5	1	6	1	2	5	2	5	1	3		5	1	3	2	2	0	6	0	2	0
14	TCHR SEX	5	1	2	1	2	1	0	1	1	1	2	2	0	2		1	0	3	1	0	0	3	0	1	0
15	TCHR QUAL	40	16	6	10	6	1	2	2	3	1	3	5	2	0		2	2	1	3	1	1	0	1	3	1
16	12 CHEM EXP	7	4	4	3	2	2	3	1	1	2	2	3	2	2		1	2	3	2	1	2	1	1	1	0
17	13 CHEM EXP	24	13	9	3	2	5	11	4	2	7	2	2	2	3		3	2	6	3	2	1	1	1	1	0
18	PD LENGTH	20	11	1	6	5	3	2	1	3	3	3	0	2	1		1	1	4	1	1	2	1	0	0	0
19	PREPS	37	14	10	7	5	3	3	2	6	3	4	4	2	4		2	2	3	2	0	2	6	2	4	0
20	12 CLASSES	25	17	7	5	4	8	6	4	1	5	2	2	1	2		1	1	3	3	1	4	1	2	1	0
21	13 CLASSES	36	10	10	10	8	4	4	3	4	3	1	3	2	1		2	4	2	3	1	3	0	1	2	0
22	TCHG PDS	20	11	5	6	5	7	6	3	4	5	2	2	2	4		2	1	3	2	1	3	1	1	2	1
23	TCHG TIME	20	11	3	6	5	3	4	1	4	3	2	2	2	2		1	1	3	2	1	2	2	1	2	1
24	TOT PUPILS	28	17	12	4	3	8	5	8	2	4	2	4	2	3		2	2	2	2	1	2	1	3	3	0
25	CHEM PD/WK	5	2	10	3	2	1	10	2	2	9	0	2	2	9		1	1	0	2	0	0	0	3	2	0
26	CLASS SIZE	60	17	15	10	7	3	4	2	5	4	2	3	5	3		2	2	4	2	2	1	1	2	4	1
27	TCHR	66	45	15	22	4	19	10	14	3	9	2	6	2	7		6	2	6	1	4	7	2	3	5	1

* Unexplained Final Group
† Explained Final Group

○ Split made on this variable

Verbal, Theoretic-Immediate 12 x SATO Mathematics, Educational Plans 2 x SATO Mathematics.

As indicated in Table 15, none of the variables other than those actually used as splitters was of potential use in splitting the groups. The total variance explained by each important variable in the run is given in Table 16.

TABLE 16

AID ANALYSIS FOR RUN NO. 3 — CONTRIBUTION OF IMPORTANT VARIABLES TO OTAC TOTAL SCORE VARIANCE

Variable	Percent of Variance Explained
SATO MATH	27.72
SATO TV	10.01
TCHR	2.37
T-I 12	.77
ED PLANS 2	.69
P-T 12	.52
Total	42.08

In this run 42.08% of the variance of OTAC total scores is explained by six variables. None of the remaining variables used in this run reduced the unexplained sum of squares by as much as 0.5%.

The nature of the variable labelled "teacher" needs some comment. When the six teachers in Group No. 8 and Group No. 16 (Figure 6) were identified, it was discovered that the six teachers came from only three schools. None of these three schools had other teachers of chemistry on the staff.

Although the splits in Run No. 3 were made on the basis of the teacher variable, it is quite possible that the school rather than the teacher was the actual splitting variable. Since two splits were made on the basis of this variable, the possibility exists that one split was made on the basis of the teacher variable and the other split made on the basis of the school variable. A further possibility is that the splitting variable was both the teacher and the school; in other words a teacher x school interaction could be responsible for the splits observed.

Because the nature of the variable itself is uncertain, it was decided to rename the variable. "School environment" was the name selected and is the name that will be used in the remainder of this report.

Since a number of teacher and school characteristics were collected as data and used as input variables in the AID runs and none of these characteristics proved to be of actual or potential use in splitting groups, one cannot in this study identify which teacher or school characteristic or characteristics are related to the explainable OTAC total score variance. All that can be concluded is that the important characteristic or characteristics cannot be one or more of those variables investigated in the present study.

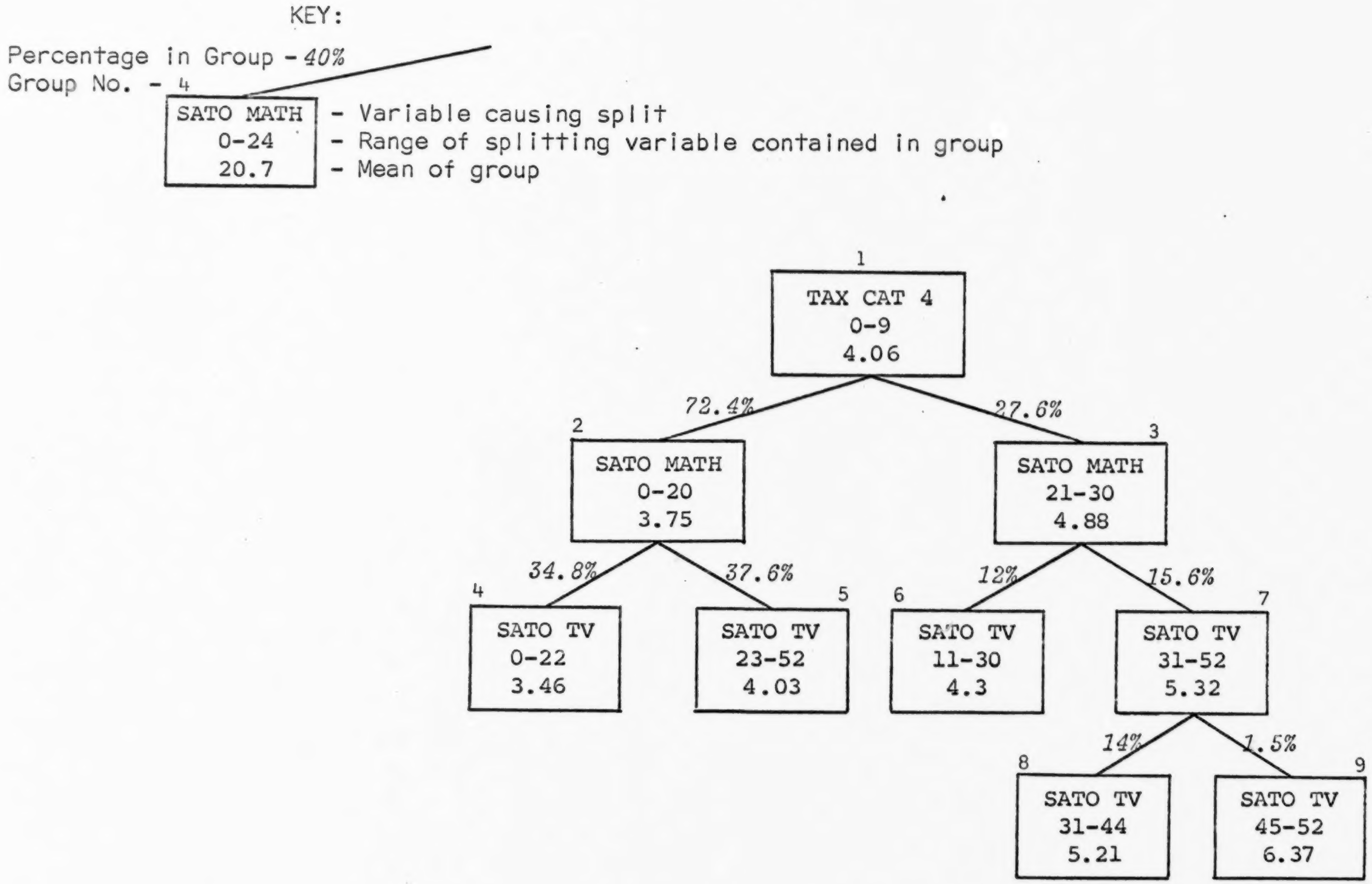

KEY:
Percentage in Group - 40%
Group No. - 4
SATO MATH
0-24
20.7
- Variable causing split
- Range of splitting variable contained in group
- Mean of group
1
TAX CAT 4
0-9
4.06
72.4%
27.6%
2
SATO MATH
0-20
3.75
3
SATO MATH
21-30
4.88
34.8%
37.6%
12%
15.6%
4
SATO TV
0-22
3.46
5
SATO TV
23-52
4.03
6
SATO TV
11-30
4.3
7
SATO TV
31-52
5.32
14%
1.5%
8
SATO TV
31-44
5.21
9
SATO TV
45-52
6.37

FIG. 7 - AID Tree for OTAC Taxonomy Category 4.00 Scores - Run No. 4

TABLE 17

AID ANALYSIS FOR RUN NO. 4 -
OTAC TAXONOMY CATEGORY 4.00 SCORE
(Between-Group Sum of Squares for Each Predictor at Each Stage)[a]

Variable		Group Number								
		1	2	5*	3	4*	7	8*	6*	9*
2	SATO TV	6	①	0	②	15	㊸	9	23	4
3	SATO MATH	⑥	1	0	1	7	39	20	21	20
4	P-T 4	1	0	0	0	11	15	12	7	17
5	P-I 4	0	0	0	0	10	12	18	10	6
6	P-A 4	0	0	0	0	4	2	1	4	2
7	T-I 4	1	0	0	0	23	28	16	3	16
8	T-A 4	1	0	0	0	8	19	10	2	15
9	A-I 4	0	0	0	0	2	4	0	2	13
10	SEX	0	0	0	0	0	4	4	6	5
11	OCCF	0	0	0	0	9	11	16	4	4
12	OCCM	0	0	0	0	2	8	7	9	3
13	OCCS	0	0	0	0	12	13	8	12	12
14	REPEATING	0	0	0	0	0	18	13	8	7
15	ATTITUDE	0	0	0	0	2	6	5	13	16
16	LANGUAGE	0	0	0	0	8	1	0	4	12
17	ED PLANS 1	3	1 ◄	0	0	25	16	9	27	16
18	SCHOOL TYPE	0	0	0	0	3	27	24	3	8
19	TEXT	0	0	0	0	0	4	4	0	0

* Unexplained Final Group

○ Split made on this variable

◄ Points to next best splitting variable

[a] For groups 1,2,5, & 3 entries are BSS × 10^2;
for groups 4,7,8,6 & 9 entries are BSS × 1.

Run No. 4

The low estimate of internal consistency obtained for the Analysis subtest has already been noted. Because of the unusual nature of the Taxonomy Category 4.00 subtest scores, it was decided to compare an AID analysis of these with the analysis of OTAC total scores. The independent variables used were the same as those of Run No. 1.

It is clearly shown in Figure 7 that mutually exclusive subgroups were formed by the action of only two of the 18 explanatory variables, viz: SATO Mathematics and SATO Verbal. No interaction between these two variables is indicated.

As indicated in Table 17, Educational Plans 1 would be almost as good a splitter as SATO Total Verbal for group 2. The total variance explained by each important variable in the run is shown in Table 18.

TABLE 18

AID ANALYSIS FOR RUN NO. 4 —
CONTRIBUTION OF IMPORTANT VARIABLES
TO OTAC SUBTEST 4 SCORE VARIANCE

Variable	Percent of Variance Explained
SATO MATH	7.69
SATO TV	4.48
Total	12.17

Thus 12.17% of the variance of OTAC total scores is explained by two variables. None of the remaining variables used in this run reduced the unexplained sum of squares by as much as 0.5%.

Run No. 5

It was decided to compare the AID analysis of final marks in chemistry with the first analysis of OTAC total scores. The independent variables used were the same as those in Run No. 1.

As shown in Figure 8, again a number of mutually exclusive subgroups are formed by the action of relatively few explanatory variables: only five out of the 18 independent variables used in this run account for the explainable variance in final chemistry marks. The following interactions are indicated: Educational Plans 1 x SATO Mathematics, SATO Total Verbal x SATO Mathematics, Repeating x Educational Plans 1, Educational Plans 1 x SATO Total Verbal, Theoretic-Immediate 4 x Educational Plans 1.

As indicated in Table 19, the variable Prudent-Theoretic 4 would be almost as good a splitter as SATO Total Verbal for group 4. The total variance explained by each important variable in the run is given in Table 20.

Thus 21.51% of the variance of final chemistry marks is explained by five variables. None of the remaining variables used in this run reduced the unexplained sum of squares by as much as 0.5%.

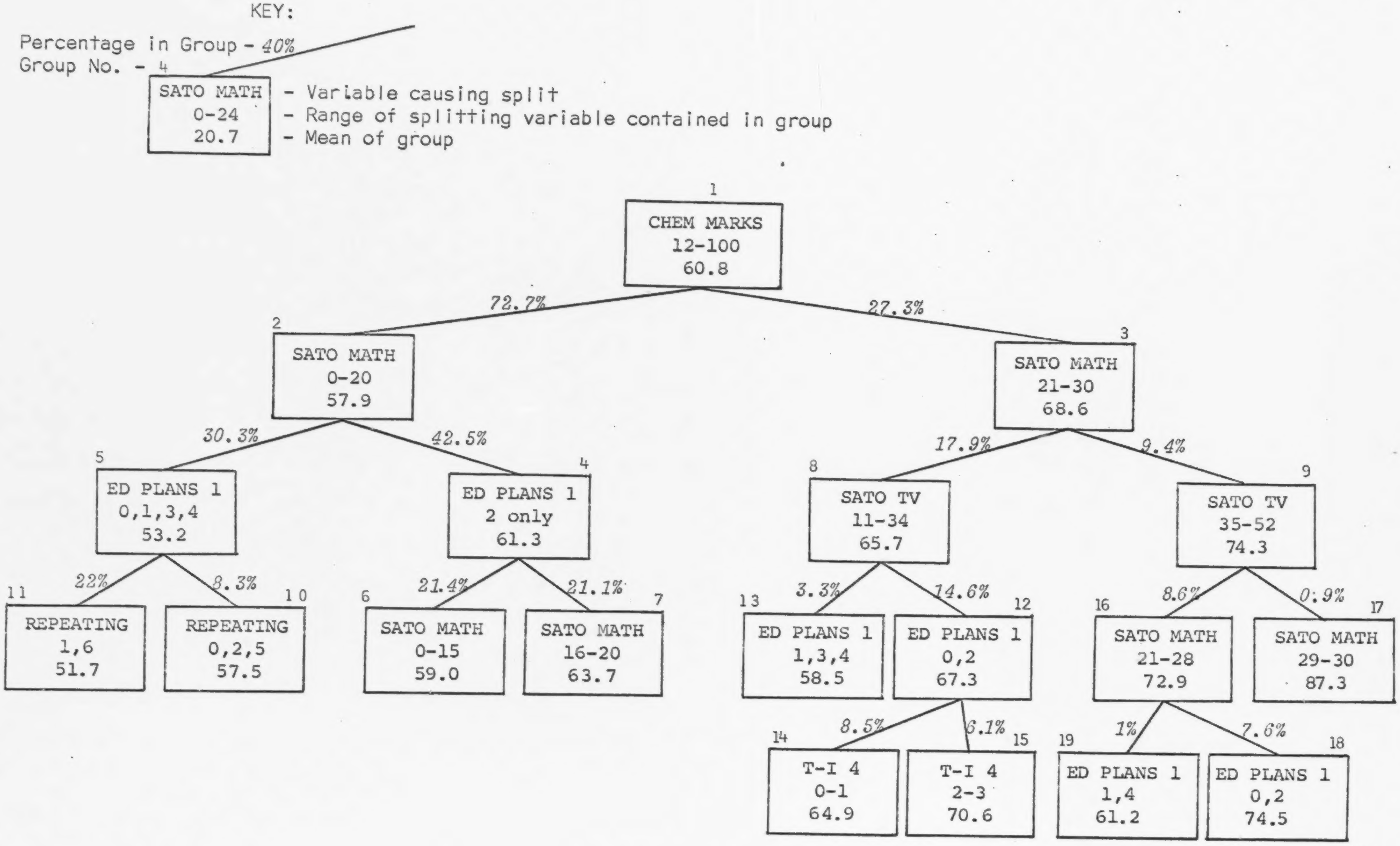

KEY:
Percentage in Group - 40%
Group No. - 4
SATO MATH
0-24
20.7
- Variable causing split
- Range of splitting variable contained in group
- Mean of group
1
CHEM MARKS
12-100
60.8
72.7%
27.3%
2
SATO MATH
0-20
57.9
3
SATO MATH
21-30
68.6
30.3%
42.5%
17.9%
9.4%
5
ED PLANS 1
0,1,3,4
53.2
4
ED PLANS 1
2 only
61.3
8
SATO TV
11-34
65.7
9
SATO TV
35-52
74.3
22%
8.3%
21.4%
21.1%
3.3%
14.6%
8.6%
0.9%
11
REPEATING
1,6
51.7
10
REPEATING
0,2,5
57.5
6
SATO MATH
0-15
59.0
7
SATO MATH
16-20
63.7
13
ED PLANS 1
1,3,4
58.5
12
ED PLANS 1
0,2
67.3
16
SATO MATH
21-28
72.9
17
SATO MATH
29-30
87.3
8.5%
6.1%
1%
7.6%
14
T-I 4
0-1
64.9
15
T-I 4
2-3
70.6
19
ED PLANS 1
1,4
61.2
18
ED PLANS 1
0,2
74.5

FIG. 8 - AID Tree for Chemistry Marks - Run No. 5

TABLE 19

AID ANALYSIS FOR RUN NO. 5 -
CHEMISTRY MARKS
(Between-Group Sum of Squares for Each Predictor at Each Stage) × 10^2

Variable		Group Number																		
		1	2	4	3	5	11*	8	7*	6*	12	9	16	17¶	14*	10*	15*	18*	19¶	13*
2	SATO VERB	97	28	7	(26)	5	5	7	4	4	5	2	1		4	5	3	1		3
3	SATO MATH	(131)	33	(14)	13	8	5	7	1	3	6	(9)	2		4	1	1	1		1
4	P-T 4	24	9	11◄	8	2	2	8	4	5	4	3	1		3	1	0	1		5
5	P-I 4	7	3	0	3	1	2	3	1	4	1	1	1		1	1	1	1		3
6	P-A 4	12	10	2	3	2	4	1	1	1	1	1	1		2	0	1	1		0
7	T-I 4	43	16	9	17	4	5	8	3	5	(7)	7	5		1	4	0	3		1
8	T-A 4	27	12	5	4	5	5	3	3	2	3	0	0		1	1	1	1		1
9	A-I 4	5	1	1	3	1	1	0	1	0	0	2	1		1	0	1	1		1
10	SEX	4	2	0	2	1	1	0	1	0	0	1	2		0	2	0	1		0
11	OCCF	6	4	2	2	3	5	1	3	0	1	2	2		2	2	2	1		1
12	OCCM	2	3	0	3	4	3	1	0	0	0	3	4		1	2	2	1		1
13	OCCS	28	14	2	4	2	1	3	1	2	2	2	2		2	2	0	1		0
14	REPEATING	4	2	2	3	(12)	1	0	0	1	0	2	2		0	2	1	3		2
15	ATTITUDE	21	15	4	12	6	6	5	6	1	4	4	4		3	2	1	2		1
16	LANGUAGE	3	3	2	1	9	0	0	2	0	0	0	1		1	2	1	0		2
17	ED PLANS 1	130	(66)	-	24	3	2	(12)	-	-	0	8	(9)		0	2	0	0		1
18	SCHOOL TYPE	9	3	1	12	2	1	6	0	2	4	3	3		5	1	0	1		0
19	TEXT	15	9	6	3	5	3	3	4	1	3	1	1		5	1	0	0		0

* Unexplained Final Group
¶ Small Final Group

○ Split made on this variable
◄ Points to next best splitting variable

TABLE 20

AID ANALYSIS FOR RUN NO. 5— CONTRIBUTION OF IMPORTANT VARIABLES TO CHEMISTRY MARKS VARIANCE

Variable	Percent of Variance Explained
SATO MATH	11.62
SATO TV	1.97
ED PLANS 1.	6.52
REPEATING	.90
T-I 4	.50
Total	21.51

Summary and Discussion of AID Analysis Results

In each of the five AID runs, not more than six variables accounted for the explainable variance. As might be expected, SATO Mathematics and Total Verbal scores accounted for most of the explainable variance, regardless of whether OTAC Total score, Taxonomy Category 4.00 subtest score or final chemistry marks was the dependent variable investigated. Immediate educational plans was an effective variable in explaining the variance of OTAC Total score and final chemistry marks. A noteworthy discovery was the importance of one or other of the Theoretic-Immediate scales of the Inventory of Choices in explaining OTAC Total score variance (and chemistry final marks variance). Other variables (Prudent-Theoretic 4- and 12-point scales, future educational plans, best subject) which accounted for some of the OTAC Total variance each reduced the unexplained sum of squares by less than 1% of the total; the single exception to this statement was a variable associated with the school environment—a variable which could not be identified from the data collected in this study. The variables sex and repeating were shown to have limited potential utility as splitters.

None of the remaining variables was of any practical use in reducing the unexplained variance of OTAC Total scores.

Rather remarkable was the finding that, of the input variables expected to explain Taxonomy Category 4.00 subtest score variance, only the two SATO scores proved important.

A number of interactions were indicated in the AID analysis but those located in the "twigs" of their respective AID trees are not likely to contribute considerably to the explainable variance of the dependent variable; the SATO Total Verbal x SATO Mathematics interaction is an exception to this statement. Some of the variables which interact with the SATO combination are studied in the pattern analyses which follow.

Examination of the AID tree diagrams (Figures 4 - 6) reveals a finding not reported in the research reviewed in Chapter II; namely, that the Theoretic-Immediate and Prudent-Theoretic scales (and immediate educational plans as well) do not act with equal efficiency across the whole group of students, but act most effectively in certain ranges of scholastic aptitude. For OTAC Total scores the Theoretic-Immediate scales (whether 4-, 9- or 12-point) discriminate best in the score ranges of 20-26 for SATO Mathematics and 29-52 for SATO Total Verbal. The 4-point Prudent-Theoretic scales operates most effectively in the score ranges of 13-19 for SATO Mathematics and 25-51 for SATO Total Verbal. Unlike its 4-point counterpart, the 12-point Prudent-Theoretic scale discriminates best in the score range of 20-30 for SATO Mathematics and 20-28 for SATO Total Verbal. In Figures 4 and 5 it is seen that immediate educational plans discriminates best in the score range of 0-19 for SATO Mathematics and 0-24 for SATO Total Verbal.

The discriminations on OTAC Total score effected by the variables ED PLANS 1, P-T 4 and T-I 4 range from 3.0 to 4.1 raw score points, that is, approximately from .37 to .50 standard deviations.

Since the object of the present investigation is to examine patterns of achievement in chemistry (OTAC scores) and their relationship to personal, attitudinal, and environmental factors (which have been segregated by the AID analysis), it is now feasible to select for further consideration those variables which may affect significantly the patterns of achievement as measured by OTAC subtest scores. It is reasonable to discard all independent variables shown to be unimportant in the AID analysis; those variables of limited potential usefulness also may be eliminated since they are overshadowed by the important variables detected in the AID analysis. Of the important variables we may conveniently confine our attention to those which were selected in Run No. 1. Run No. 1 yielded the highest overall percentage of variance explained by factors which could be clearly identified. The factor of relative achievement (underachievement, normal achievement, overachievement) could not be included in the AID analysis, but is considered in the analyses which follow. The variables thus selected for further study comprise SATO Mathematics, SATO Total Verbal, Educational Plans 1, Theoretic-Immediate 4, Prudent-Theoretic 4, and relative achievement; the influence of these variables upon achievement profiles will now be considered.

Pattern Analysis

Preliminary Treatment of the Data

In the previous chapter it was proposed to use residual scores rather than raw scores to form profiles. The residual score is defined as the observed score minus the predicted score. One of the educational hypotheses was that OTAC scores would show substantial correlation with SATO Total Verbal and Mathematics scores; the data shown in Table 21 indicate that such is generally the case.

Regression equations may be calculated for OTAC Total scores and *Taxonomy* subtest scores using the formulas provided by Garrett (1953, p.392). Details of the procedure are given in Appendix N.

TABLE 21

CORRELATIONS BETWEEN OTAC AND SATO SCORES

Test	SATO Total Verbal	SATO Mathematics
OTAC Total Score	.5306	.5834
Taxonomy Category 1.00	.4393	.4339
Taxonomy Category 2.00	.4331	.5255
Taxonomy Category 3.00	.4506	.5654
Taxonomy Category 4.00	.3236	.2895
SATO Total Verbal	--	.4907

The regression equations take the following form when two predictor variables are used:

$$\hat{X}_1 = b_2X_2 + b_3X_3 + C$$

where $\hat{X}_1$ is the predicted score

X_2 is the first predictor variable

X_3 is the second predictor variable

b_2 and b_3 are regression coefficients for X_2 and X_3 respectively

C is a constant.

A program was written by the present writer to compute regression coefficients, constants, predicted scores, and residual scores for each of the 2,248 students for whom SATO scores were available. The following are the regression equations obtained:

For OTAC Total score : $\hat{X} = 0.3145\,V + 0.6085\,M + 6.6539$

Taxonomy Category 1.00: $\hat{X} = 0.1254\,V + 0.1770\,M + 3.4525$

Taxonomy Category 2.00: $\hat{X} = 0.0644\,V + 0.1683\,M + 0.7224$

Taxonomy Category 3.00: $\hat{X} = 0.0710\,V + 0.2066\,M + 0.8285$

Taxonomy Category 4.00: $\hat{X} = 0.0537\,V + 0.0566\,M + 1.6509$

in each case $\hat{X}$ is the predicted score

V is the SATO Total Verbal score

M is the SATO Mathematics score.

Once all residual scores had been computed, they were arranged in frequency distributions, normalized, and standardized by means of a computer program written by the present writer. The normalization section of the program used two function subprograms developed by Dr. J. C. Ogilvie of the Department of Mathematics, University of Toronto, to calculate the standard score, given the proportion of the total area under the normal probability curve. The normalized standardized residual scores were expressed as T scores which have a mean of 50 and a standard deviation of 10 (Guilford, 1956, pp.494-500).

The final step in preparing the data for Haggard's method of pattern analysis was to stabilize the normalized standardized residual scores for the four subtests by dividing these scores by their respective standard errors of measurement. The standard error of measurement is given by the formula

$$\text{s.e.}_{\text{meas}} = \sigma_t\sqrt{1 - r_{tt}}$$

where $\text{s.e.}_{\text{meas}}$ is the standard error of measurement

σ_t is the standard deviation of the test or subtest

r_{tt} is the test-retest reliability of the test or subtest.

In Haggard's method of pattern analysis r_{tt} must not be estimated from the data being analyzed (Haggard, 1958, pp. 104-108); the value of r_{tt} used in the present case is the test-retest reliability obtained by correlating the OTAC Total scores and subtest scores obtained in the test-retest administration which was conducted separately from the main OTAC administration.

Since the scores to be stabilized are residuals, that is, scores with the SATO Total Verbal and Mathematics components removed, and the test-retest standard deviations and reliabilities were computed from raw scores with these components not removed, adjusted values of the standard deviations, the test-retest reliabilities, and hence the standard errors of measurement, ought to be employed in stabilizing the residual score.[5] The procedure for adjusting the test-retest reliabilities, and standard errors of measurement is described in Appendix N. Adjusted values were used in the profile analysis; these values are listed in Table 22.

---

[5] The writer is indebted to Dr. Ross E. Traub for suggesting the appropriate adjustment.

TABLE 22

COMPARISON OF ORIGINAL AND ADJUSTED STANDARD ERRORS OF MEASUREMENT

Test	Original Value Based on Raw Scores	T-score Value, Not Adjusted for Presence of SATO Components	T-score Value, Adjusted for Presence of SATO Components
OTAC Total score	3.440	--	--
Taxonomy Category 1.00	1.942	5.550	5.815
Taxonomy Category 2.00	1.345	5.788	6.986
Taxonomy Category 3.00	1.385	5.339	6.906
Taxonomy Category 4.00	1.547	8.276	9.058

A computer program written by the present investigator produced the adjusted standard errors of measurement. In the pattern analysis, to be described shortly, T scores were used as input, and the division by the adjusted values of the standard error of measurement was performed as the first step.

Definitions of Overachievement, Underachievement, and Normal Achievement

Differences between obtained and predicted scores are used to gauge the relative achievement of students; for example, an individual is considered to be an underachiever when his obtained score is lower than his expected (predicted) score by a predetermined (and usually arbitrary) amount. In the present study the differences (residual scores) have been calculated and arranged in a frequency distribution; overachievers and underachievers may be considered to be those students whose residual scores lie toward the extremes of the distribution. Arbitrary dividing lines must be selected, and in the present research it was decided to consider as over- and underachievers those students whose residual scores on the total test fell outside the range $\pm\sigma$ (that is, plus or minus one standard deviation from the mean). This criterion of over- and underachievement is similar to that proposed by Thorndike (1963, p.63).

In terms of T scores, normal achievers thus have residual scores on Total OTAC falling in the range 40-60; underachievers have scores below 40 and overachievers have scores above 60. Since the residual scores are normalized, overachievers comprise approximately the uppermost one-sixth of the distribution and underachievers the lowest one-sixth.

Methods of Forming Groups for Analysis

With the use of residual scores in the profile analysis, the immediate educational plans of the student and the Theoretic-Immediate and Prudent-Theoretic scales of the Inventory of Choices are left as variables useful in identifying groups of students. The AID program isolated groups in which these variables acted most selectively. Two methods of forming groups for profile analysis were used.

④

One method was to analyze groups selected by the AID program and broken into subgroups on the basis of the three variables mentioned. The other method was to stratify all students as under-, over-, or normal achievers and to examine, within these strata, groups based on the three variables referred to above.

The Decision Process

The statistical hypotheses have been discussed in Chapter III (pp.111-113). The decision process followed in conducting the analysis may be summarized conveniently in the flow-chart of Figure 9; in the analysis of each subgroup the procedure followed is that indicated in the upper half of Figure 9.

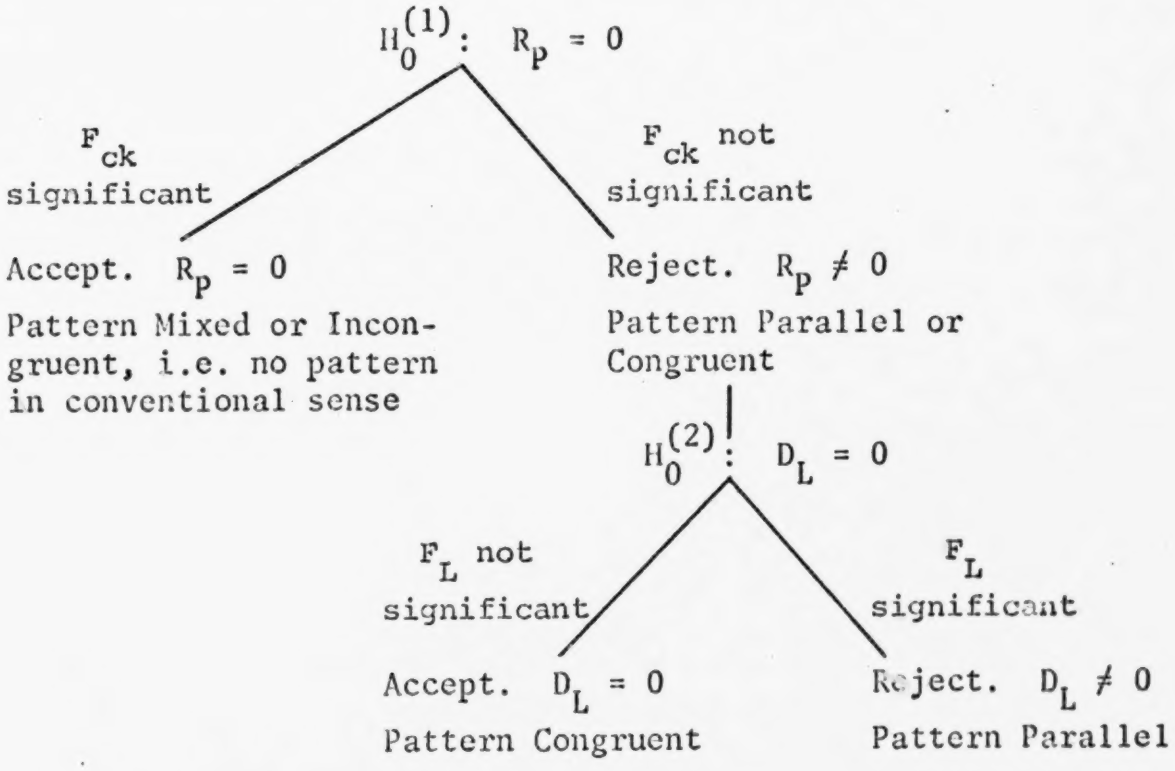
Where two or more patterns were identified, the profiles forming the patterns were pooled and analyzed as indicated in the lower half of Figure 9.

The Program

Two decks of cards containing stabilized scores were created; one deck was sorted to retrieve AID groups No. 4, 11, and 12 of Run No. 1 (see Figure 4), while the other deck was sorted into relative achievement groups. Each group was re-sorted into subgroups on the basis of immediate educational plans, Prudent-Theoretic score, or Theoretic-Immediate score.

A two-way analysis of variance forms the essential

Q.1. Does one group have a pattern? If yes, what kind?


$H_0^{(1)}: R_p = 0$
F_{ck} significant
F_{ck} not significant
Accept. $R_p = 0$
Pattern Mixed or Incongruent, i.e. no pattern in conventional sense
Reject. $R_p \neq 0$
Pattern Parallel or Congruent
$H_0^{(2)}: D_L = 0$
F_L not significant
F_L significant
Accept. $D_L = 0$
Pattern Congruent
Reject. $D_L \neq 0$
Pattern Parallel

Q.2. If patterns exist among two or more groups, are patterns different or same? If same, what kind?

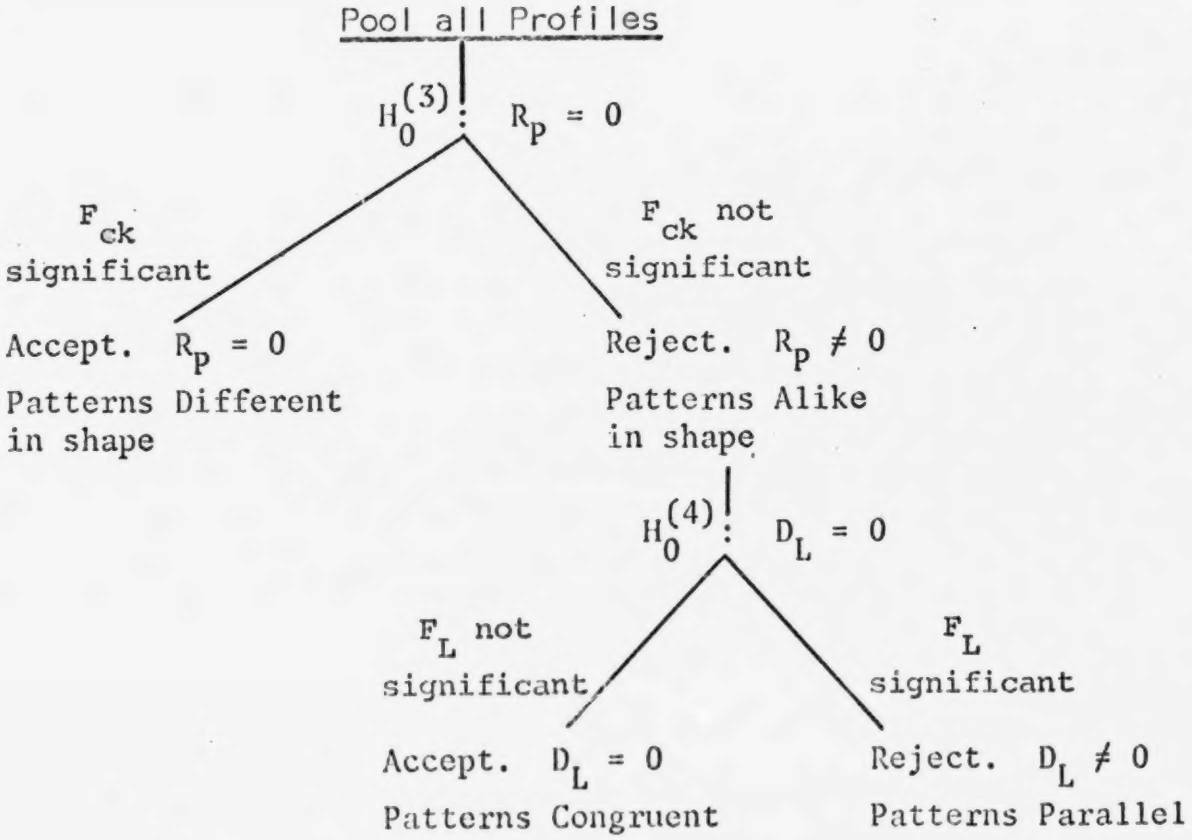

Pool all Profiles
$H_0^{(3)}: R_p = 0$
F_{ck} significant
F_{ck} not significant
Accept. $R_p = 0$
Patterns Different in shape
Reject. $R_p \neq 0$
Patterns Alike in shape
$H_0^{(4)}: D_L = 0$
F_L not significant
F_L significant
Accept. $D_L = 0$
Patterns Congruent
Reject. $D_L \neq 0$
Patterns Parallel

Fig. 9. Decision Process Flowchart

part of Haggard's method of pattern analysis; a computer program was written by the present writer, incorporating the ANØVA Library Tape Subroutine written for the IBM 7094 by Dr. J. C. Ogilvie. Since individual students' mean scores were not required for pattern analysis, the calculation of these was suppressed. Identification of non-significant interaction F's was facilitated by the fact that these were numerically equal to the interaction mean squares, since stabilized scores were used as input.

The pattern analysis program made one iteration on each subgroup and then made iterations on combinations of subgroups. All four categories of the Taxonomy were included in the first profile analysis; a further analysis was then made on the four categories considered three at a time.

Results

The pattern analysis computer output provides an analysis of variance table for each group and subgroup studied. Where congruent or parallel patterns emerge across a complete group, the relevant analysis of variance table will be found in Appendix I. In this section check charts and summary tables are used as a condensed form of the information found in Appendix I.

Groups Selected by the AID Program

Inspection of the check chart (Table 23) shows that, where all four subtests are considered, no congruent or parallel patterns emerge across all subdivisions of any of the groups selected by the program.

TABLE 23

CHECK CHART OF SIGNIFICANT PROFILE SIMILARITIES
FOR GROUPS SELECTED BY THE AID PROGRAM

Taxonomy Configuration	AID Group	Subgroup 1 (Subgroup with lower mean score on OTAC Total)	Subgroup 2 (Subgroup with higher mean score on OTAC Total)	Entire Group (Subgroups 1 & 2 combined)
1 2 3 4	T-I 4			
	P-T 4			
	ED PLANS 1			
1 2 3	T-I 4			
	P-T 4			
	ED PLANS 1			
1 2 4	T-I 4			
	P-T 4			
	ED PLANS 1			
1 3 4	T-I 4			
	P-T 4			
	ED PLANS 1			
2 3 4	T-I 4			
	P-T 4			
	ED PLANS 1	X[a]	X	X

[a] R_p significant at .05 level.

When profiles consisting of only three subtests are considered, in only one case do congruent or parallel patterns emerge across a whole group. In this case, profiles consisting of Categories 2.00, 3.00, and 4.00 show congruent or parallel patterns in the AID group selected on the basis of immediate educational plans. Analyses for the two subgroups alone and in combination are presented in Table 24.

TABLE 24

SUMMARY OF PATTERN ANALYSES FOR AID
GROUP ED PLANS 1

Group	n	F_{ck}	p	R_P	p	F_L	p
1. Not intending to complete Grade 13	431	1.21		.83	.05	2.07	.001
2. Intending to complete Grade 13	419	1.20		.83	.05	2.37	.001
3. Groups 1 and 2 combined	850	1.20		.83	.05	2.26	.001

In group 1 the interaction F (F_{ck}) is not significant and leads to rejection of the first null hypothesis for that group. It is concluded that the profiles in group 1 form a definite pattern. The F for the differences in level (F_L) is significant, allowing a rejection of the second null hypothesis for group 1. Thus it is concluded that the pattern for group 1 is parallel rather than congruent. Similar results lead to similar conclusions for group 2.

When groups 1 and 2 are pooled, the interaction F is found to be not significant, leading to a rejection of the third null hypothesis; it is then concluded that the two patterns are of similar shape. The F for the differences in level is significant, allowing a rejection of the fourth null hypothesis; thus it is concluded that the two patterns of profiles are parallel rather than congruent.

It is seen therefore that two patterns of similar shape but differing in level emerge from that portion of the total group which is most sensitive to selection on the basis of immediate educational plans; a noteworthy characteristic of this group is that it is that portion of the total OTAC group whose SATO Total Verbal score is below 25 and whose SATO Mathematics score is below 20.

The patterns of the two groups are plotted in Figure 10.

Groups Stratified on Relative Achievement

In contrast to the pattern analysis just described, which was performed on selected groups of students isolated by the AID Program, the following analyses involved all students for whom residual scores could be calculated. The factor Educational Plans 1, which has only two effective categories, was used to dichotomize overachievers, normal achievers, and underachievers separately. The Theoretic-Immediate and Prudent-Theoretic scales (each consisting of four categories) were used to subdivide each classification of relative achievement. The resulting unit categories were analyzed separately by the pattern analysis program. In the case of the Inventory of Choices scales, a dichotomy was effected later in the program by pooling appropriate

- 6 -

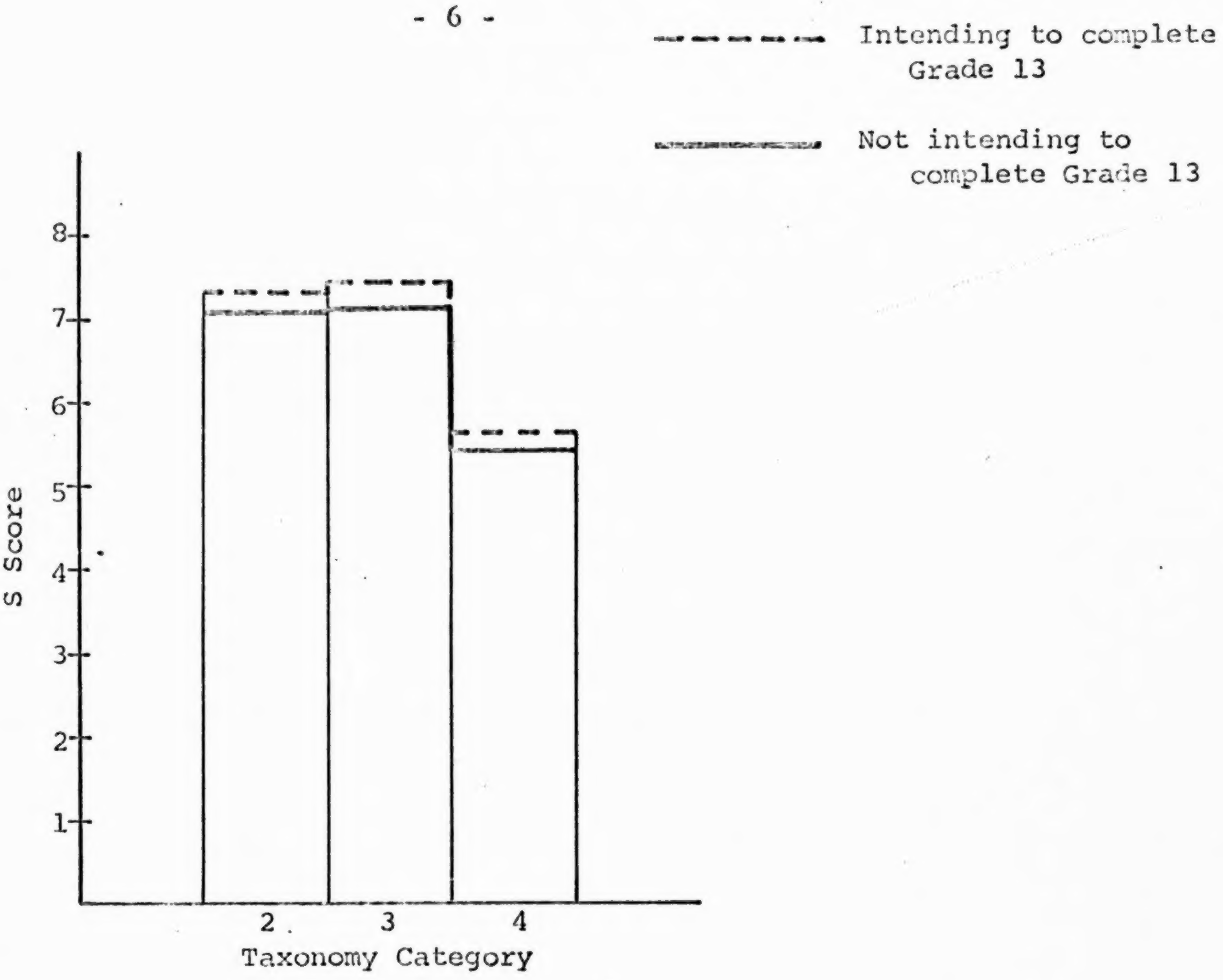


FIG. 10 - Patterns of AID groups with Differing Immediate Educational Plans

categories in pairs; the combined categories were then pattern analyzed. All four Taxonomy categories were included in the pattern analysis and then combinations of three of the four categories were studied.

Reference to the check charts (Table 25, 26, and 27) shows that patterns did not emerge throughout an achievement classification when all four Taxonomy subtests were analyzed together. When combinations of four subtests taken three at a time were analyzed, patterns emerged across all categories only when Taxonomy Categories 2.00, 3.00, and 4.00 constituted subtests in the profiles, and then only for underachievers and overachievers in the Theoretic-Immediate subdivisions, and for overachievers in the Prudent-Theoretic subdivisions. The immediate educational plans subdivisions contained no congruent or parallel patterns across both categories in any Taxonomy configuration.

In the pattern analysis tables which follow, in every case profiles and patterns consist only of subtest scores in Categories 2.00, 3.00 and 4.00 of the Taxonomy.

Table 28 and 29 contain the analyses of the underachieving and overachieving groups of students falling into each unit category of the Theoretic-Immediate Scale.

In none of the eight unit categories is the interaction F significant; the first null hypothesis is therefore rejected in each case. The F for the difference in level

TABLE 25

CHECK CHART OF SIGNIFICANT PROFILE SIMILARITIES FOR THEORETIC-IMMEDIATE GROUPS

Taxonomy Configuration	Group	Theoretic-Immediate Score					
		4-point Scale				Dichotomized Scale	
		0	1	2	3	0 & 1	2 & 3
1 2 3 4	Underachievers						
	Achievers						
	Overachievers	X[a]	X		X		
1 2 3	Underachievers				X		
	Achievers						
	Overachievers	X			X	X	
1 2 4	Underachievers		X	X			
	Achievers						
	Overachievers	X	X		X	X	
1 3 4	Underachievers			X			
	Achievers						
	Overachievers		X	X	X		
2 3 4	Underachievers	X	X	X	X	X	X
	Achievers			X			
	Overachievers	X	X	X	X	X	X

[a] R_P significant at .05 level.

TABLE 26

CHECK CHART OF SIGNIFICANT PROFILE SIMILARITIES FOR PRUDENT-THEORETIC GROUPS

Taxonomy Configuration	Group	Prudent-Theoretic Score					
		4-point Scale				Dichotomized Scale	
		0	1	2	3	0 & 1	2 & 3
1 2 3 4	Underachievers	X[a]					
	Achievers						
	Overachievers	X	X		X	X	
1 2 3	Underachievers	X					
	Achievers						
	Overachievers	X	X			X	
1 2 4	Underachievers	X	X			X	
	Achievers						
	Overachievers	X	X		X	X	
1 3 4	Underachievers	X					
	Achievers						
	Overachievers	X	X		X	X	
2 3 4	Underachievers	X	X	X		X	
	Achievers	X					
	Overachievers	X	X	X	X	X	X

[a] R_P significant at .05 level.

TABLE 27

CHECK CHART OF SIGNIFICANT PROFILE SIMILARITIES FOR IMMEDIATE EDUCATIONAL PLANS GROUPS

Taxonomy Configuration	Group	ED PLANS 1 Subgroup	
		Not Intending to Complete Grade 13	Intend to Complete Grade 13
1 2 3 4	Underachievers	X[a]	
	Achievers		
	Overachievers		X
1 2 3	Underachievers		
	Achievers		
	Overachievers		X
1 2 4	Underachievers	X	
	Achievers		
	Overachievers		X
1 3 4	Underachievers		
	Achievers		
	Overachievers		X
2 3 4	Underachievers	X	
	Achievers		
	Overachievers		X

[a] R_P significant at .05 level.

is not significant for any of the eight unit categories; the second null hypothesis is therefore rejected in each case.

TABLE 28

SUMMARY OF PATTERN ANALYSES OF UNDERACHIEVERS

Theoretic-Immediate Score	n	F_{ck}	p	R_P	p	F_L	p
0	56	1.15		.87	.05	1.22	
1	112	1.13		.88	.05	1.18	
2	59	1.10		.91	.05	.86	
3	21	1.26		.80	.05	.91	

TABLE 29

SUMMARY OF PATTERN ANALYSES OF OVERACHIEVERS

Theoretic-Immediate Score	n	F_{ck}	p	R_P	p	F_L	p
0	26	1.31		.77	.05	.70	
1	93	1.06		.95	.05	.96	
2	100	1.22		.82	.05	.99	
3	58	1.02		.98	.05	1.04	

Thus it is seen that for underachievers and overachievers each score category of the Theoretic-Immediate scale contains a congruent pattern of profiles made up of scores on Categories 2.00, 3.00, and 4.00 of the Taxonomy.

When the Theoretic-Immediate scales were dichotomized, the results were as shown in Table 30.

TABLE 30

SUMMARY OF PATTERN ANALYSES—DICHOTOMIZED THEORETIC-IMMEDIATE SCALE

Group	Theoretic-Immediate Scores	n	F_{ck}	p	R_p	p	F_L	p
Underachievers	0 & 1	168	1.14		.88	.05	1.19	
	2 & 3	80	1.13		.89	.05	.86	
Overachievers	0 & 1	119	1.13		.88	.05	.91	
	2 & 3	158	1.14		.88	.05	1.01	

In all four combined categories each interaction F is non-significant, leading to rejection of the third null hypothesis for each paired group. The F for the differences in levels in each case is also non-significant, leading to acceptance of the fourth null hypothesis for each paired group.

Thus it is seen that for both underachievers and overachievers each half of the Theoretic-Immediate scale contains a congruent pattern of profiles made up of scores on Categories 2.00, 3.00, and 4.00 of the Taxonomy.

It now seems reasonable to ask, "Do all underachievers having a Theoretic-Immediate score have the same pattern?" and "Do all overachievers having a Theoretic-Immediate score have the same pattern?" To answer these questions the profiles of these underachievers were combined and analyzed; treated similarly were the profiles of these overachievers. The results are presented in Table 31.

TABLE 31

SUMMARY OF PATTERN ANALYSES OF ALL UNDERACHIEVERS AND ALL OVERACHIEVERS HAVING A THEORETIC-IMMEDIATE SCORE

Group	Theoretic-Immediate Score	n	F_{ck}	p	R_P	p	F_L	p
Underachievers	0, 1, 2, or 3	248	1.15		.87	.05	1.08	
Overachievers	0, 1, 2, or 3	277	1.13		.88	.05	.96	

Inspection of Table 31 shows that both interaction F's are not significant, leading to a rejection of the third null hypothesis. The F's for the differences in levels likewise are not significant, leading to acceptance of the fourth null hypothesis.

It is thus concluded that for Categories 2.00, 3.00, and 4,00, the profiles of all underachievers have a congruent pattern regardless of Theoretic-Immediate score, and that the corresponding profiles of all overachievers have a congruent pattern regardless of Theoretic-Immediate score.

The underachievers and overachievers for whom Theoretic-Immediate scores were obtained are subsets of the underachievers and overachievers in the sample studied. Do these subsets have congruent patterns by virtue of their having a Theoretic-Immediate score, or is the congruence a characteristic of all underachievers and overachievers? To answer this question a pattern analysis was performed on all underachievers in the sample and also on all overachievers in the sample. The results of these two analyses are found in Table 32.

TABLE 32

SUMMARY OF PATTERN ANALYSES OF ALL UNDERACHIEVERS AND ALL OVERACHIEVERS

Group	n	F_{ck}	p	R_P	p	F_L	p
All Underachievers	312	1.20		.83	.05	1.16	
All Overachievers	337	1.15		.87	.05	0.93	

Examination of Table 32 shows that both interaction F's are not significant, allowing a rejection of the third null hypothesis in each case. The F's for the differences in levels likewise are not significant leading to acceptance of the fourth null hypothesis for each group.

It is thus concluded that all underachievers, regardless of whether or not they could be classified along the Theoretic-Immediate continuum, have profiles which form a congruent pattern. The same conclusion is drawn for all overachievers in the sample.

The question, "Do the patterns of underachievers and overachievers have the same shape?" now presents itself. To answer this question the profiles of all underachievers and overachievers were pooled and analyzed; Table 33 presents the results of that analysis.

TABLE 33

SUMMARY OF PATTERN ANALYSIS OF ALL UNDERACHIEVERS AND OVERACHIEVERS COMBINED

n	F_{ck}	p	R_P	p	F_L	p
649	1.27	.025			6.59	.001

The interaction F is significant at the .025 level leading to an acceptance of the third null hypothesis. Thus it is concluded that while all underachievers have similar profiles and all overachievers have similar profiles, the shape of the underachievers' pattern is not the same as that of the overachievers. Similar conclusions were reached for the Theoretic-Immediate subsets of underachievers and overachievers (Table I-21, Appendix I). The patterns of underachievers and overachievers are plotted in Figure 11.

Since it has been shown that underachievers or overachievers who could be classified along the Theoretic-Immediate continuum have the pattern characteristics of all underachievers or all overachievers, it hardly seems necessary to repeat here the summary analyses and accompanying comments for overachievers for whom a Prudent-Theoretic score was obtained; Appendix I has the relevant tables for these groups. It should be noted however, (Table 26), that the Prudent-Theoretic subset of underachievers does not retain the congruence of pattern possessed by all underachievers. The classification of underachievers along the Prudent-Theoretic continuum results in profiles not forming congruent or parallel patterns in some categories.

Dichotomizing the underachievers or overachievers according to immediate educational plans (Table 27) reveals a lack of congruent or parallel patterns across the resulting groups.

The consistently congruent patterns appearing in all Theoretic-Immediate categories for both underachievers and overachievers suggests a further investigation. One other way of combining unit categories exists. It may be asked, "Do the underachievers and overachievers having a similar Theoretic-Immediate score possess patterns of similar shape?" To answer this question the appropriate groups of underachievers and overachievers were pooled and analyzed. The results are presented in Table 34.

In the work that follows the present writer uses "non-normal achievers" and "combinations of underachievers and overachievers" synonymously.

TABLE 34

SUMMARY OF PATTERN ANALYSES OF COMBINATIONS OF
UNDERACHIEVERS AND OVERACHIEVERS HAVING
THE SAME THEORETIC-IMMEDIATE SCORE

Theoretic-Immediate Score	n	F_{ck}	p	R_P	p	F_L	p
0	82	1.22		.82	.05	5.62	.001
1	205	1.22		.82	.05	6.55	.001
2	159	1.25	.05			5.71	.001
3	79	1.16		.86	.05	5.18	.001

In three of the four groups the interaction F is not significant, leading to rejection of the third null hypothesis; in these groups the F's for differences in level are, of course, significant, leading to rejection of the fourth null hypothesis. For the group with a Theoretic-Immediate score of 2 the third null hypothesis is accepted.

It is thus seen that when underachievers and overachievers in any Theoretic-Immediate category are compared for each extreme group in the Theoretic-Immediate continuum the two patterns are of the same shape although they differ in level. One moderate group of overachievers and underachievers combined has patterns of a common shape but the other group has not.[6] The patterns of each extreme group of underachievers and overachievers are plotted in Figure 12.

It may also be considered whether both extreme groups of underachievers and overachievers combined have patterns of the same shape. Combining underachievers and overachievers from Theoretic-Immediate score categories 0 and 3, and analyzing this combined group yields the results shown in Table 35.

---

[6] The two moderate groups of non-normal achievers thus cannot be studied together and are not considered further in this analysis.

In Table 35 the interaction F is significant at the .05 level. For the combined group the third null hypothesis may be accepted.

TABLE 35

SUMMARY OF PATTERN ANALYSES OF THE COMBINATION OF UNDERACHIEVERS AND OVERACHIEVERS WITH EXTREME THEORETIC-IMMEDIATE SCORES

Group	Theoretic-Immediate Score	n	F_{ck}	p	R_p	p	F_L	p
Extreme	0 & 3	161	1.26	.05			6.61	.001

It is thus concluded that the two extreme groups of non-normal achievers do not have patterns of similar shape.

Thus far the analysis of patterns has been concerned only with the first four hypotheses as delineated in Figure 9. Testing of the fifth, sixth and seventh hypotheses will be undertaken now for those patterns which have been found congruent or parallel.

Differences Between Means

Congruent or parallel patterns emerged from the ED PLANS 1 group selected by the AID program, from underachiever and overachiever groups, and from combined underachiever and overachiever (or "non-normal achiever") groups which had Theoretic-Immediate scores of 0, 1, or 3. Since in none of these cases did Taxonomy Category 1.00 subtest scores contribute to consistent patterns, the fifth null hypothesis

$$H_0^{(5)}: \ M_1 = M_2 = M_3 = M_4 \quad \text{now reduces to}$$

$$H_0^{(5)}: \ M_2 = M_3 = M_4$$

For the same reason the sixth null hypothesis now becomes

$$H_0^{(6)}: \ M_i = M_j \ ; \quad i \neq j, \ \begin{array}{l} i = 2, 3 \\ j = i + 1, \ldots, 4 \end{array}$$

TABLE 36

AMONG-SUBTESTS F'S FOR GROUPS HAVING CONGRUENT OR PARALLEL PATTERNS ACROSS CATEGORIES 2.00, 3.00, AND 4.00

df = 2;182

Group	Subgroup	F_c	p
AID ED PLANS 1	Not intending to complete Grade 13	314.82	.001
	Intending to complete Grade 13	343.71	.001
All Underachievers		109.48	.001
All Overachievers		487.51	.001
Underachievers and Overachievers Combined	Highly Immediate	32.67	.001
	Highly Theoretic	89.92	.001

The data in Table 36 reveal that for each group the fifth null hypothesis may be rejected at the .001 level. The significance of the differences between pairs of means may now be tested. For this purpose the studentized range statistic (Winer, 1962, pp.77-85) is suitable.

The studentized range statistic, q, is defined by

$$q = \frac{M_i - M_j}{\sqrt{RMS/n}}$$

where M_i and M_j are treatment means

RMS is the residual (error) mean square

and n is the number of observations in each treatment group.

Since the residual mean square is unity in Haggard's method of pattern analysis, the above equation may be rewritten as

$$M_i - M_j = q\sqrt{1/n}$$

Using $q_{.99}$ (k,f) to designate the 99th percentile point on the q distribution, with k = number of treatments and f = degrees of freedom for RMS,

$$\text{if} \left| M_i - M_j \right| \geqq q_{.99}\ (k,f)\ \sqrt{1/n}$$

the difference between the two means is significant at the .01 level. A similar statement may be made for the .05 level of significance using $q_{.95}$. The results of testing the sixth hypothesis for each pair of means for each group are presented in Table 37. In this instance k = 3, f = 182, $q_{.99}$ (3,182) = 4.12, and $q_{.95}$ (3,182) = 3.31.

TABLE 37

DIFFERENCES BETWEEN PAIRS OF MEANS IN GROUPS HAVING CONGRUENT OR PARALLEL PATTERNS ACROSS CATEGORIES 2.00, 3.00, AND 4.00

Group	Subgroup	n	Pair of Means		Difference of Means		
			i	j	Critical Values $q\sqrt{1/n}$		Observed $\lvert M_i - M_j \rvert$
					.01	.05	
AID ED PLANS 1	Not intending to complete Grade 13	431	2	3	.1985	.1594	.0401
			2	4			1.6061**
			3	4			1.6462**
	Intending to complete Grade 13	419	2	3	.2013	.1617	.1264
			2	4			1.6523**
			3	4			1.7787**
All Underachievers		312	2	3	.2332	.1874	.0853
			2	4			1.0789**
			3	4			1.1642**
All Overachievers		336	2	3	.2248	.1806	.1809*
			2	4			2.1336**
			3	4			2.3145**
Non-normal Achievers (Underachievers and Overachievers Combined)	Highly Immediate	82	2	3	.4550	.3655	.2599
			2	4			1.3143**
			3	4			1.0544**
	Highly Theoretic	79	2	3	.4635	.3724	.3741*
			2	4			1.7794**
			3	4			2.1535**

**significant at the .01 level

* significant at the .05 level

Inspection of Table 37 reveals that the sixth null hypothesis may be rejected at the .01 level of significance for all pairs of means which include Category 4.00, and at the .05 level for two pairs of means embracing Categories 2.00 and 3.00. It is thus concluded that significant differences in means exist between Categories 2.00 and 4.00 and between Categories 3.00 and 4.00. No significant differences exist between the means of Categories 2.00 and 3.00 except for overachievers and for highly Theoretic non-normal achievers.

The proposed alternative hypotheses to $H^{(5)}$ may now be restated as:

$$H_P^{(5)}: \; M_2 > M_3 > M_4$$

$$H_I^{(5)}: \; M_2 > M_3 > M_4$$

$$H_T^{(5)}: \; M_2 < M_3 < M_4$$

None of these alternative hypotheses can be accepted. Groups identified as Prudent were seen to have consistently congruent patterns only as overachievers. Groups identified as Immediate or Theoretic likewise demonstrated consistently congruent patterns only as underachievers or overachievers. Reference to Figures 11 and 12 and Table 37 shows that in all cases at least one step between subtests either was in a direction opposite to that hypothesized or was not significant.

It does not seem feasible to formulate, a priori, meaningful alternate fifth hypotheses for the AID ED PLANS 1 groups or for the underachiever-overachiever groups.

The seventh null hypothesis of the form

$$H_0^{(7)}: \bar{M}_A = \bar{M}_B$$

can be tested by computing F_L for each pair of patterns concerned. Suitable alternative hypotheses are the following:

$$H_{EP}^{(7)}: \bar{M}_{13} > \bar{M}_{\text{not } 13}$$

$$H_{UA-OA}^{(7)}: \bar{M}_{OA} > \bar{M}_{UA}$$

$$H_{TI}^{(7)}: \bar{M}_T > \bar{M}_I \text{ (for non-normal achievers only)}$$

Values for F_L are reported in Tables 24, 33 and 35. In each case F_L is significant at the .001 level and leads to rejection of the corresponding seventh null hypothesis. The alternative hypotheses are accepted. The first two alternative hypotheses have been included for the sake of completeness: the difference in level between the two ED PLANS 1 groups has been established by the AID program; the difference in level between underachievers and overachievers follows by definition.

The two proposed alternate hypotheses

$$H_{PT}^{(7)}: \bar{M}_T > \bar{M}_P$$

$$H_{PI}^{(7)}: \bar{M}_P > \bar{M}_I$$

cannot be considered further since it is pointless to test their corresponding null hypotheses.[7,8]

---

[7] All Prudent-Theoretic classifications of overachievers were shown to have the same pattern (Table I-34, Appendix I). Reference to Table 26 shows that some Prudent-Theoretic classifications of underachievers did not form congruent or parallel patterns and therefore the desired comparisons of grand means of patterns cannot be made.

[8] No attempt was made to investigate patterns involving Prudent-Immediate classifications since this variable was shown by the AID analysis to be relatively ineffective.

It is seen that the overall or pattern mean of the AID group intending to complete Grade 13 is significantly higher (0.2489 S score units) than the overall (grand) mean of the group not intending to complete Grade 13. The overall mean of the highly Theoretic non-normal achievers is significantly higher (1.2815 S score units) than the pattern mean of their highly Immediate counterparts. As must be expected, the overall mean of overachievers is significantly higher (2.7219 S score units) than the grand mean of the underachievers.

The difference in pattern means of the two AID ED PLANS 1 groups is significant but small. On the other hand the difference in pattern means for the two non-normal achiever groups is approximately one half the difference found between the pattern means of overachievers and underachievers.

The common pattern of overachievers shows a peak at Category 3.00; the same effect is found in that group of non-normal achievers which is highly Theoretic. In any other group there is no significant difference in the means of Categories 2.00 and 3.00.

All patterns in Figures 10, 11, and 12 show a substantial drop from Category 3.00 to 4.00. While the relatively large standard error of measurement of subtest 4 must account for a considerable part of this effect, it is important to note that for overachievers (Figure 11 and Table 37) the drop is approximately twice that of the underachievers. In a similar manner, for highly Theoretic underachievers and overachievers (Figure 12 and Table 37) the drop is about twice that of their highly Immediate counterparts. One should note also the similarities in the patterns of overachievers and highly Theoretic non-normal achievers, on the one hand, and the similarities in the patterns of underachievers and the highly Immediate non-normal achievers, on the other.

Comparison of Findings with Those of Similar Studies

While the present study appears to be the only one reported to date in which the Taxonomy was used to form achievement profiles, a number of findings which emerge in the present research merit comparison with those of related studies which concern (a) tests constructed according to the Taxonomy, (b) factors related to the achievement of specific objectives of high school chemistry, and (c) the Inventory of Choices.

Taxonomy-type tests

McFall (1964) and McGuire (1963b) obtained low correlations between subtests which were constructed to measure specific cognitive objectives, and considered that these low correlations indicated that the subtests were measuring different abilities. In the present study the subtest intercorrelations were not generally as low as those reported by McFall or McGuire, possibly because the present writer attempted to equalize difficulty over subtests; the three OTAC subtests whose intercorrelations were more or less uniform (.54-58) were also alike in difficulty (.42-48). The correlations of the Category 4.00 subtest to the other three subtests were noticeably lower (.39-41) and that subtest was also more difficult (.33) than were the others.

The above findings contrast sharply with those of Schmitt _et al_. (1966) where all intercorrelations were higher than .80. The work of K. Anderson (1949), and Porter and Anderson (1959) may be considered taxonomic in intention; however, their categories are more broadly defined than those of the _Taxonomy_ and on analysis each category is seen to include more than one _Taxonomy_ level. The subtest intercorrelations in both these studies generally were somewhat higher than observed in the present study, but not as high as those reported by Schmitt _et al_.

The tendency of the subtest correlation matrix to approach a simplex was noted by McGuire. Thomas (1965) reported the emergence of a simplex, as does the present study. Stoker and Kropp (1964) obtained a number of simplexes in their analyses. Generally, the simplex model is approached by matrices in which the taxonomic level does not exceed Category 4.00.

Low subtest reliabilities seem to be characteristic of _Taxonomy_ subtests so far reported. For Categories 1.00-3.00, Herron (1966) reported reliabilities of about the same magnitude as those given in the present study. Herron found Category 4.00 reliabilities to be the lowest, as did the present investigator. Herron also reported that the reliabilities of June Anderson's (1964) taxonomy-type test were low. Herron's overall test reliability was approximately the same as found in the present study. In contrast, K. Anderson's quasi-taxonomy subtests generally had much higher reliabilities, (.76-.89) although one subtest (acquisition of scientific attitudes) had a much lower reliability than the others. Porter and Anderson's reliabilities were also higher than those of the present study.

Ayers (1966) mentioned the desirability of longer tests; the present study has shown that longer subtests would likely result in substantial improvement in test reliability.

The studies of relationships of specific cognitive objectives to intelligence or scholastic aptitude show some interesting results. McFall found that his knowledge subtest correlated more highly with IQ than did the subtest measuring higher cognitive abilities. Thomas (1965) found a decreasing correlation of IQ with increasing taxonomic level, and an increasing correlation of reasoning scores with taxonomic level.

The present study found no such monotonic relationship, possibly because verbal and mathematical components of scholastic aptitude were measured separately;

however, Category 4.00 showed much lower correlations with the two SATO scores than did the three other categories.

Anderson (1949) found higher correlations of all subtests with IQ except for the scientific attitude subtest where no significant correlation was observed.

Porter and Anderson (1959) found all subtests correlated equally well with IQ.

Factors Related to High School Chemistry Achievement

The findings of the study by Anderson (1949,1950) afford some interesting comparisons with the present study. In the present study the relationship of many variables to chemistry achievement was investigated by means of the AID program. Both studies found that chemistry achievement was significantly related to the educational plans of the student, but not to the sex or age of the student. The number of teacher preparations generally did not show a relationship to student chemistry performance in either study. On the other hand, Anderson found larger class size, experience of the teacher, use of a laboratory manual, and the number of college Chemistry or science credits held by the teacher to be positively related to student chemistry achievement.[9]

---

[9] In this comparison the factor of teacher qualification was approximated by using the Ontario Secondary School Teachers' Federation categories assigned to teachers in Ontario. Category 1 represents the lowest certifiable qualification and category 4 the highest.

Such relationships were not observed in the present study - the AID program found that such variables made unimportant contributions to the explainable variance of OTAC total scores when other (better) explanatory variables were available.

The Inventory of Choices

The present study reveals some findings concerning the Inventory of Choices and its relationship to chemistry achievement. The main findings are mentioned here and other findings are reported in Appendix K.

In the comments that follow, it must be kept in mind that most of the studies reviewed used grades or marks assigned by teachers as the criterion variable; the data collected in the present study include both teacher marks and objective test scores. The comments concerning the present study refer only to results involving the objective test scores. Teacher-assigned marks are treated in Appendixes L and M.

The usefulness of Inventory of Choices scales as correlates of chemistry achievement was investigated by the AID program. Two scales (Theoretic-Immediate and Prudent-Theoretic) were shown to make important contributions to the explainable variance of the criterion in combination with scholastic aptitude. The remaining Inventory of Choices scales were shown not to function well in combination with the two important scales or with other (better) predictors of chemistry achievement.

The findings of the present study support the findings of Edwards and Wilson (1958b; 1959b): for example, the average chemistry score difference observed by Edwards and Wilson (1959b) was about .34 σ for the Prudent-Theoretic scale, when IQ and pretest scores were held constant, whereas the present study showed a difference of .37 σ for the Prudent-Theoretic scale and .50 σ for the Theoretic-Immediate scale when SATO Mathematics and SATO Total Verbal contributions to score variance were accounted for separately. However, in the present study the effectiveness of the Prudent-Theoretic scale was not as large as expected and the Prudent-Immediate scale (Edwards and Wilson, 1961) did not compare well with the Prudent-Theoretic or Theoretic-Immediate scales as discriminators.

The findings of Maykovich (1966) were supported in part by the findings of the present study. Maykovich observed that, from the ninth grade to the beginning of the twelfth grade, a migration away from the Theoretic pole took place. Reference to Appendix K shows that, during the Grade 12 year, significant changes away from the Theoretic orientation were made on the Theoretic-Aesthetic and Theoretic-Immediate scales by the Grade 12 group studied in this research. The change on the Prudent-Theoretic 4-point scale was also significant.

Maykovich reported that Prudent students showed marked superiority over Theoretic students in all subjects except chemistry and geometry. The AID program revealed that in certain ranges of scholastic aptitude, Theoretic students were significantly superior to Prudent and Immediate students in chemistry achievement, as measured by OTAC. An additional finding was that in those ranges of scholastic aptitude the Prudent-Theoretic and Theoretic-Immediate scales were more effective than any other explanatory variables tried. The selectivity of the Prudent-Theoretic and Theoretic-Immediate scales in certain ranges of scholastic aptitude has not been reported previously.

If one assumes that the migration away from the Theoretic pole has taken place over the high school years for the sample in this study, as well as in Maykovich's sample, the greater discriminatory effect of the Theoretic-Immediate scale is readily explained: the group at the Theoretic pole is enriched by the emigration of less successful students, thus heightening the contrast in achievement between the Theoretic and Immediate groups. This finding is consonant with the theoretical structure of the Inventory of Choices: the Theoretic-Immediate scale distinguishes between both the social versus non-social orientation and the deliberative versus non-deliberative orientation.

Summary and Discussion of the Findings

The data presented in Table 4 show that, in the sample tested, large variations occurred in the attainment of cognitive objectives. In Taxonomy Categories 1.00, 2.00, and 3.00, scores obtained by students covered the entire possible range of scores for each subject; in Category 4.00, a 12-item subtest and the most difficult of the four, scores from 0 to 10 were obtained. The relative dispersion of each subtest has been noted in Appendix G; most scattering of scores occurred in Category 2.00, with less in Category 3.00, still less in Category 4.00, and least in Category 1.00.

Correlations between OTAC scores and SATO scores (Table 21) show that scores in Categories 2.00 and 3.00 are more highly related to mathematical ability than to verbal ability, whereas the reverse is true for scores in Category 4.00; scores in Category 1.00 are related about equally to both mathematical and verbal abilities as measured by SATO. Correlations between SATO scores and scores in Categories 1.00, 2.00, and 3.00 are substantial, while the correlations between SATO and Category 4.00 scores are low.

Reference to Appendix M shows that correlations between OTAC Total score and final chemistry mark or final average mark were substantial, as was the correlation between OTAC Total score and SATO scores. Correlations between OTAC subtest scores and SATO scores or teacher-assigned marks were not as low as expected, except in the case of Category 4.00 scores.

The cognitive abilities measured by the subtests of OTAC show moderate correlation with each other (Table 8).

Categories 1.00, 2.00, and 3.00 as a group show substantial correlation with each other; on the other hand the correlations between Category 4.00 and other subtests are noticeably lower.

The observations noted above suggest that the set of abilities measured by Category 4.00 is somewhat anomalous, whereas the other three Categories share properties in common to a considerable extent. The much lower reliability (Table 4) and larger standard error of measurement of Category 4.00 (Table 22) also point to its singularity.

The relationship between chemistry achievement, as measured by OTAC, and a large number of personal, attitudinal, and environmental factors was explored by means of the Automatic Interaction Detector (AID) program. The major finding in this area was that many of the variables commonly thought to influence chemistry achievement made no important contribution (that is, in excess of 0.5%) to the explainable variance of the total chemistry score.

The bulk of OTAC variance explained was contributed to by SATO Mathematics and SATO Total Verbal scores, with the former outweighing the latter by a factor of more than 2.5. The remainder of the explainable variance was largely accounted for by the school environment, immediate educational plans of the student, and score on a Theoretic-Immediate scale; small amounts of variance were explained by Prudent-Theoretic scale scores, future educational plans, and best subject. Although a number of variables characteristic of the teacher and school were investigated, none of these accounted for the variance explained by the school environment. Lacking more precise information, one might for the time being attribute this variation to the individuality of the teacher, the school, or both teacher and school.

More than half of the observed variance remains unexplained; such variance may be attributed to variables which were not investigated, to individual differences, and to "noise".

The peculiar nature of Category 4.00 is further attested to by its analysis by the AID program. No variables other than SATO Mathematics and SATO Total Verbal account for any important segment of the explainable variance. The fact that only 12% of the variance is explained by the SATO scores discourages one from concluding that the abilities tested by Category 4.00 are merely scholastic aptitude in its various forms.

A number of interactions were detected by the AID program. The most frequent interaction seen in the AID trees was SATO Verbal x SATO Mathematics, although this interaction was not detected in the analysis of Category 4.00. Higher order interactions invariably involved SATO Total Verbal x SATO Mathematics. Some interactions which involved either or both SATO variables were investigated indirectly in the pattern analysis where residual scores were used. School environment and future educational plans interacted with other variables but were not considered in the pattern analysis. The interactions involving future educational plans occurred on the "twigs" of the AID tree and likely would not contribute much to the explainable variance. The school environment variable's elusive quality discourages complicated analytical procedures; it should suffice to record that SATO Total Verbal x School Environment and SATO Mathematics x School Environment interactions were observed. The interactions indicate that a complex situation remains once the simple factors have been extracted.

The AID analyses indicate that few variables are available for study in relation to patterns of achievement as defined in the present study. The application of Haggard's pattern analytic techniques show that congruent or parallel patterns of achievement profiles do not exist across all four Categories of the Taxonomy for any groups, including groups of under-, over-, or normal achievers for whom Theoretic-Immediate or Prudent-Theoretic scores were available, or for the same groups dichotomized with respect to their immediate educational plans. The same is true for those students who were isolated most selectively on these variables by the AID program.

Only when profiles were restricted to Categories 2.00, 3.00, and 4.00 did congruent or parallel patterns emerge, and then only for over- and underachievers, and for the groups singled out by the AID program on the basis of immediate educational plans.

The two groups separated by the AID program exhibited parallel patterns which differed in level by small but significant amounts. Students planning to complete Grade 13 had a pattern of slightly higher level than that of students not intending to complete Grade 13. A noteworthy feature of these groups with parallel patterns is their below-average verbal ability and moderate mathematical ability. When groups were formed on the basis of relative achievement, pattern analysis revealed that normal achievers did not have congruent or parallel patterns across Categories 2.00, 3.00, and 4.00, either when considered as a single group or when grouped according to Theoretic-Immediate scores, Prudent-Theoretic scores, or immediate educational plans. On the other hand, congruent or parallel patterns were found amongst overachievers and underachievers when these were analyzed with reference to their Theoretic-Immediate scores, and amongst overachievers when these were grouped according to their Prudent-Theoretic scores. All underachievers had the same congruent pattern regardless of Theoretic-Immediate score. All overachievers had the same congruent pattern regardless of Theoretic-Immediate score or Prudent-Theoretic score. However, the patterns of underachievers and overachievers were not the same in shape (and, of course, not the same in level); overachievers showed a significant increase in score from Category 2.00 to Category 3.00 while underachievers did not. In addition, the overachievers show a drop in score from

Category 2.00 or 3.00 to Category 4.00 which is about twice the same drop for underachievers. When overachievers and underachievers of similar Theoretic-Immediate score were compared, common patterns were discovered in those groups of non-normal achievers whose position on the Theoretic-Immediate scale is extreme: highly Theoretic persons show a common parallel pattern as do highly Immediate persons, although the patterns differ in shape for each group. The highly Theoretic non-normal achievers' pattern is higher in level than that of the highly Immediate non-normal achievers and in addition shows a significant increase in score from Category 2.00 to Category 3.00. While the highly Immediate persons showed a drop in score from Category 2.00 or 3.00 to 4.00, the corresponding drop for highly Theoretic non-normal achievers is about twice as much. The highly Immediate group pattern fits a hypothesized pattern ($H_I^{(5)}$) more closely than any other pattern which emerged.

The similarity in shape of the overachievers' and highly Theoretic non-normal achievers' patterns is noticeable; underachievers and highly Immediate non-normal achievers likewise have patterns of similar shape. These similarities suggest a common factor of interest in chemistry achievement. One explanation that might be put forth is that highly Theoretic students are those who are intrinsically motivated toward chemistry achievement; overachievers are likely to be highly motivated extrinsically.

On the other hand, highly Immediate students by definition do not have intrinsic motivation toward chemistry (or any other academic subject), while underachievers as a group are probably underachieving in chemistry for extrinsic reasons.

Motivation does not explain why overachievers or non-normal highly Theoretic students make better scores in Category 3.00 than in Category 2.00. The present writer suspects that the tasks required by Category 3.00 items receive more emphasis in the chemistry classroom than that received by Category 2.00 tasks; for example, most numerical problems are found in Category 3.00. Overachievers or highly Theoretic students would likely concentrate on such tasks, having noted the relative emphasis given to these in the classroom.

The same argument may be advanced for the differential achievement observed in Category 4.00. The drop to Category 4.00 is partly the result of the subtest's comparatively large standard error of measurement, but this reason cannot account for the differences in drop observed between the two groups being compared. Overachievers may not be able to achieve as well in this type of mental activity as in the activities required for Categories 2.00 and 3.00. Underachievers are likely to do poorly in all categories, and Category 4.00 achievement may not suffer by comparison to as great an extent.

It is this investigator's impression that the tasks required for Category 4.00 are not emphasized in the introductory course which has been taught in Ontario classrooms. If this impression is true, then for highly Theoretic non-normal achievers the tasks required may represent those which are quite unfamiliar to them. Highly Immediate persons are not likely to be interested in such tasks (even if they are taught in the classroom), and their achievement level may reflect scholastic aptitude more than that of the highly Theoretic persons; the result is likely to be a reduction in the contrast between Categories 3.00 and 4.00.

It is important that the above explanation be considered to apply to only a portion of the students classed as underachievers or overachievers. For the prediction of academic achievement, ability is but one kind of necessary information; what is left after ability has been used as a predictor is unexplained variation, much of which might be accounted for by predictive factors not used in the present study. Some such factors likely would be the student's perceived importance of the test, his interest in the problems presented by the particular test administered, his susceptibility to anxiety, his degree of conformity to parental expectations of success, and so forth.

In spite of the foregoing speculations, the nature of the abilities represented in Category 4.00 remains to be explained. The AID run showed that scholastic aptitude explains only a small part of the explainable variance;

what factors explain the remainder remains to be investigated. The possibility that Category 4.00 may be multidimensional should be seriously considered. Also to be considered is the possibility that, in spite of satisfactory item statistics, the items in Category 4.00 were poorly constructed for the purpose of measuring representatively the abilities subsumed under the Taxonomy heading "Analysis".

Why profiles involving only Categories 2.00, 3.00, and 4.00 give rise to coherent patterns is not answered by the present study. One advantage of having clearly defined patterns is that, given a student's score on a particular subtest represented in the pattern, and given the group to which the student belongs, one is able to predict the student's score on other subtests. The findings show that students with similar achievement profiles in Categories 2.00, 3.00, and 4.00 may be expected to vary widely in their achievement in Category 1.00. The findings lead one to conclude that specific chemistry knowledge is not predictable on the same bases as non-knowledge objectives, or that common achievement trends in Categories 2.00, 3.00, and 4.00 are not reflected in the acquisition of knowledge.

While Category 4.00 seems to possess a measure of singularity not found in the three lower categories, at least some common achievement trends may be discerned in the profiles consisting of Taxonomy Categories 2.00, 3.00, and

4.00. On the other hand, Category 1.00 possesses a measure of unpredictability not found in the three higher categories when profiles are considered. Kropp, Stoker, and Bashaw's (1966) suggestion that Category 1.00 may be multidimensional is worth exploring in the light of these findings.

It seems that, depending on the specific research purpose an investigator has in mind, useful groupings of the _Taxonomy_ categories would be "1.00 versus not 1.00", (as used by McFall, 1964), "not 4.00 versus 4.00", or for analysis of profiles, 1.00, 2.00 and 3.00, and 4.00.

Why normal achievers' scores do not form coherent patterns across the score range of the variables investigated is difficult to answer. While normal achievers, by definition, form large groups, the group size alone cannot explain the relatively large interaction F's observed: some small groups had large interaction F's while some large groups had non-significant interaction F's.

The absence of stronger relationships than those observed in the present study may lead one to conclude that OTAC itself is insensitive to most of the factors expected to have some bearing on achievement in chemistry, and is not sufficiently sensitive to measure adequately the objectives it purports to measure. This criticism is valid to some degree; while extensive cross-validation of OTAC was not attempted in the present study, the reliability of the subtests (a precondition for validity) seems low, and longer subtests would quite likely improve both the reliability and validity of each subtest. The test may appear especially insensitive to achievement in Category 4.00, but a possibility exists that skills in this cognitive area were not taught to a substantial extent in the classroom. It may also be claimed that a good chemistry test _should_ be relatively free of the influence of factors such as those which were studied in this investigation; perhaps the present writer has succeeded too well in this respect.

The categories of the _Taxonomy_ comprise wide sets of abilities and it is quite likely that considerable overlap will occur between categories. The dependence upon factual knowledge of items in higher categories also tends to prevent clear definition of the _Taxonomy_'s main subdivisions. The apparent lack of sensitivity of OTAC may thus be due, in part, to some measure of ambiguity inherent in the higher levels of the _Taxonomy_; the condensation of the _Taxonomy_ into fewer main categories carried out by McFall (1964) and Winter _et al_.(1965) supports this contention.

Lack of sensitivity may be related also to the attempt to keep difficulty constant from subtest to subtest; it is possible that item difficulty cannot be divorced from taxonomic level. However, since a good simplex was obtained, it is evident that hierarchical structure amongst the subtests existed in spite of the attempt to equalize subtest difficulty.

Since OTAC meets adequately the criteria of conventional tests and compares favorably in those characteristics empirically reported for Taxonomy-type tests, it remains debatable whether the test is generally insensitive or whether it is relatively immune to the influence of those factors which are indirectly concerned with chemistry achievement. The comments of the teachers whose students wrote OTAC, the comments of the panel of judges, and the verification of the test's scientific accuracy by two eminent chemists familiar with the teaching of high school chemistry in Ontario lead to the conclusion that the test has high curricular validity for the program of studies followed in Ontario. It must be admitted that for the purpose of analyzing profiles according to scores obtained on Taxonomy category subtests, the test may not be adequate; at present there is a scarcity of empirical evidence regarding the sensitivity required of Taxonomy-type tests to reveal differences in profile structure.

The data collected permitted a number of comparisons to be made which were not concerned with the patterns of achievement measured by OTAC but which are related to high school chemistry achievement in Ontario. Appendix L contains some results of analyses of covariance applied to achievement in the cognitive objectives measured by OTAC; Appendix M contains some remarks concerning achievement in chemistry as measured by teachers' marks, and the relationship of these marks to OTAC scores and average marks obtained by students on final examinations.

CHAPTER VI

SUMMARY AND CONCLUSIONS

Summary

The purpose of the present study was to describe the variations which occur in the attainment of cognitive objectives in high school chemistry, to identify patterns of achievement in terms of these cognitive objectives, and to investigate the relationship of achievement of these objectives and their patterns to certain personal, attitudinal, and environmental factors.

The cognitive objectives studied were restricted to Knowledge, Comprehension, Application, and Analysis as defined by the Taxonomy of Educational Objectives, Cognitive Domain. The Ontario Test of Achievement in Chemistry (OTAC), a 60-item end-of-course test designed to measure these cognitive objectives, was constructed and developed over a three-year period. Each cognitive objective was represented by a subtest; approximately 40% of the items were devoted to testing Knowledge with the remaining 60% split almost equally among Categories 2.00, 3.00, and 4.00 of the Taxonomy.

The sample consisted of 2,339 Grade 12 Chemistry students enrolled in the General Course (a college-preparatory course) in Ontario schools. Of 50 schools selected at random, 30 agreed to participate in the study. All students in the participating schools who were enrolled in Grade 12 Chemistry and in attendance on the testing day wrote the test. Most of the testees responded to a personal questionnaire and to the Inventory of Choices, a measure of attitudinal orientation developed by Edwards and Wilson.

Students' scores on the Scholastic Aptitude Test, Ontario edition (SATO) 1963-64, were retrieved from the files of the Department of Educational Research, of the Ontario College of Education, University of Toronto. Other data gathered included sex of student, educational plans and occupational aspiration, family data, some features of the home environment, some characteristics of the school environment, and characteristics of the chemistry teacher. Final marks in chemistry and average Grade 12 final marks were also obtained.

It was expected that wide variations in the attainment of the cognitive objectives would occur and that total OTAC scores would show substantial correlation to SATO scores, final chemistry marks, and average final examination marks; lower correlations of OTAC subtest scores to the above factors and between each other were also expected.[1]

It was hypothesized that distinct patterns of achievement, as measured by Taxonomy subtest scores with scholastic aptitude held constant, would emerge across groups of students falling in various classifications; it was also hypothesized that the attainment of cognitive objectives and patterns of these objectives would be related to various personal, attitudinal, and environmental factors. To illustrate, highly Prudent students would be expected to have a pattern (defined by the descending means of the subtest scores) which differed in shape from that of highly Theoretic students, and which differed in level from that of highly Immediate students. In a similar manner, other personal, attitudinal, and environmental factors might be expected to be related to patterns of achievement. Interactions among the factors were expected to arise.

---

[1] Reference to Appendix M shows that these expectations were generally fulfilled.

The analysis of the data revealed wide variation in the achievement of Knowledge, Comprehension, Application, and Analysis. The Comprehension subtest exhibited the greatest relative dispersion of scores, with Application showing less, Analysis still less, and Knowledge the least relative dispersion.

The Analysis subtest appeared to be somewhat unusual in that achievement in this area was more highly related to verbal ability than to mathematical ability, a situation not found in the other subtests; the correlation between Analysis and SATO scores was much lower than similar correlations involving the three lower Taxonomy categories. The correlations of the Analysis subtest to other subtests were much lower than the correlations observed between the other subtests. The reliability of the Analysis subtest was quite low and consequently the standard error of measurement was quite large compared to that of the other subtests.

The Automatic Interaction Detector (AID) program results revealed that a large number of personal, attitudinal, and environmental factors made no important contribution (that is, more than 0.5%) to the explainable variance of the total chemistry score. SATO Mathematics and SATO Total Verbal scores accounted for most of the variance explained, with SATO Mathematics accounting for more than 2.5 times the variance explained by the Total Verbal score. The school environment, the immediate educational plans, and the Theoretic-Immediate score of the Inventory of Choices accounted largely for the remainder of the explainable variance. Prudent-Theoretic scores, future educational plans, and best subject accounted for small amounts of the explainable variance of OTAC total scores. A number of teacher variables, such as sex, qualification, experience, and work load (but not including personality variables) failed to account for any important portion of the explainable variance.

Variables which accounted for important portions of the explainable variance of the total chemistry scores were found to act most effectively in specific ranges of aptitude. Immediate educational plans discriminated best amongst students of moderate to very low mathematics aptitude, and below-average to very low verbal aptitude. The 4-point Prudent-Theoretic scale discriminated best in the moderate to low mathematics aptitude range, and the above-average to very high verbal aptitude range. The Theoretic-Immediate scale discriminated best amongst students with above-average to very high scores in both SATO Mathematics and Total Verbal.

The unidentified school environment variable made important contributions to the explainable variance in two ranges of scholastic aptitude: (a) moderate to very low mathematics and below-average to very low verbal ability, and (b) moderate to low mathematics and average to very high verbal ability. Future educational plans likewise discriminated in two restricted score ranges, one being further restricted to students of Theoretic orientation; the best subject variable was still further restricted to students of the Theoretic subgroup intending to enter university. The Prudent-Theoretic 12-point scale did not function most effectively in the same aptitude range as the Prudent-Theoretic 4-point scale.

An AID analysis on the Analysis subtest showed only SATO Mathematics and SATO Total Verbal scores to be effective in explaining an important portion of the variance of the subtest. In this case about 88% of the variance was not accounted for, and no interaction between the two main explanatory variables was detected.

SATO Mathematics x SATO Total Verbal was the most common interaction to emerge from the AID analyses; higher order interactions invariably involved these two factors.

Some interactions involving a third important factor were investigated in the pattern analyses.

Pattern analyses were performed on residual scores which were obtained by subtracting predicted scores from obtained scores. Predicted scores were calculated from SATO Total Verbal and SATO Mathematics scores by appropriate regression equations. To meet the requirements of Haggard's method of pattern analysis, the residual subtest scores were normalized and standardized on the same scale, and then stabilized by dividing by the respective standard error of measurement adjusted for inclusion of SATO components.

Immediate educational plans, Prudent-Theoretic score, and Theoretic-Immediate score proved useful in identifying groups of students. Two methods of forming groups for analysis were used: (a) selection of groups by the AID program; (b) stratification of students on relative achievement and then formation of groups within these strata on the basis of the three identifying variables.

For groups selected by the AID program and for groups selected on the basis of over-, under-, or normal achievement, no congruent or parallel patterns emerged across all four categories of the Taxonomy. When profiles consisting of three of the four subtests were analyzed, congruent patterns emerged only when Categories 2.00, 3.00, and 4.00 were involved.

The immediate educational plans group selected by the AID program gave rise to two patterns showing small but significant differences in level; the students who intended to complete Grade 13 had a slightly higher level of achievement in all three _Taxonomy_ categories, but no significant difference in shape of the two patterns was observed.

In groups stratified according to relative achievement, congruent or parallel patterns failed to emerge across Categories 2.00, 3.00, and 4.00 for normal achievers in any grouping. Prudent-Theoretic groupings of underachievers did not give rise to congruent or parallel patterns across Categories 2.00, 3.00, and 4.00 over the complete Prudent-Theoretic scale range. Grouping according to immediate educational plans did not give rise to congruent or parallel patterns across Categories 2.00, 3.00, and 4.00 over both categories of that variable.

Groups of under- and overachievers subdivided on the basis of Theoretic-Immediate scores produced congruent patterns of profiles across Categories 2.00, 3.00, and 4.00 and over the complete Theoretic-Immediate score range. When the groups were pooled and students without Theoretic-Immediate scores included, it was found that all underachievers had the same pattern of profiles regardless of Theoretic-Immediate score and that all overachievers had a common pattern of profiles regardless of Theoretic-Immediate score. The patterns of each group in each relative achievement stratum were therefore seen to be variants of the overall pattern of that stratum. However, the two strata did not have profile patterns of the same shape (by definition the patterns must differ in level).

Comparing underachievers and overachievers of like Theoretic-Immediate score revealed that the patterns of students holding an extreme position on the Theoretic-Immediate scale were similar in shape: highly Theoretic underachievers and highly Theoretic overachievers had patterns of the same shape; highly Immediate underachievers and highly Immediate overachievers likewise had patterns of a common shape. However, the highly Theoretic patterns were not of the same shape as the highly Immediate patterns. Of course, the differences in level of the patterns of under- and overachievers were highly significant.

Tests of significance between means showed that, with two exceptions, the "step" between Categories 2.00 and 3.00 on any pattern was not significant. All other steps in all of the patterns were significant. For overachievers and highly Theoretic non-normal achievers the step from Category 2.00 to Category 3.00 was significant and the peak of the profile occurred at Category 3.00; in addition, for both of these groups the drop to Category 4.00 was approximately twice as many stabilized score units as for the underachievers or for highly Immediate non-normal achievers. The patterns of overachievers and highly Theoretic non-normal achievers thus resemble each other, as do the patterns of underachievers and highly Immediate non-normal achievers.

Conclusions

Principal Findings

The essential findings of the present study follow.

1. Large variations in the attainment of the cognitive objectives of high school chemistry appear in the sample studied. Less dispersion of achievement occurs in the Taxonomy category Knowledge than in the three higher categories tested.
2. Substantial correlations occur between OTAC[2] total scores and SATO[3] Total Verbal and SATO Mathematics scores, as well as between OTAC total scores and final marks (grades) in chemistry. Moderate correlations occur between SATO scores and OTAC subtest scores, the correlation with SATO Mathematics being generally higher than with SATO Total Verbal scores, except in the case of Analysis. The test and all subtests correlate less with final average marks than with final chemistry marks.
3. The abilities measured by the Analysis subtest appear to be somewhat anomalous. Low relationships exist between the Analysis subtest and other subtests, scholastic aptitude, final chemistry marks, and final

---

[2] Ontario Test of Achievement in Chemistry.

[3] Scholastic Aptitude Test, Ontario edition.

average marks. The Analysis subtest has a comparatively large standard error of measurement.

4. A large number of personal, attitudinal, and environmental factors have no significant relationship to achievement in chemistry as measured by OTAC. Variables which make important contributions to the explainable variance of OTAC total scores are, in descending order of importance: mathematics aptitude, verbal aptitude, school environment, immediate educational plans, Theoretic-Immediate orientation, Prudent-Theoretic orientation, future educational plans, and subject liked best. The first two characteristics listed account for most of the explainable variance; more than half the variance is unaccounted for.
5. The characteristic or characteristics of the school environment which contributed to the explainable variance of OTAC total scores were not among the many characteristics studied in this investigation, and thus could not be identified.
6. The interaction of mathematics and total verbal aptitude is prominent; higher order in interactions involve this pair of variables.
7. No congruent or parallel patterns of achievement involving all four categories of the Taxonomy appear across any group of students, including those groups selected by the AID program and those groups classified as underachievers, normal achievers, or overachievers.

8. Only when patterns consisting of Categories 2.00, 3.00 and 4.00 of the Taxonomy are considered do congruent or parallel patterns appear. Parallel patterns appear in the groups selected by the AID program on the basis of immediate educational plans; congruent patterns appear for underachievers and overachievers. Coherent patterns do not appear for normal achievers in any grouping. The overachiever pattern is different in shape from the underachiever pattern.
9. Highly Theoretic underachievers and highly Theoretic overachievers have patterns of similar shape. Highly Immediate underachievers and highly Immediate overachievers have patterns of a common shape. The combined pattern of the highly Theoretic non-normal achievers is not of the same shape as that of their highly Immediate counterparts.
10. Underachievers have a pattern in which no significant difference in score occurs between Categories 2.00 and 3.00. A significant drop to Category 4.00 is evident.
11. Overachievers have a pattern in which a significant peak occurs in Category 3.00. The drop from Category 3.00 to Category 4.00 is twice that observed for underachievers.
12. The pattern of highly Theoretic non-normal achievers resembles that of the overachievers; the pattern of highly Immediate non-normal achievers resembles that of the underachievers.

13. Common achievement trends in Categories 2.00, 3.00 and 4.00 of the Taxonomy are not reflected in any group's achievement in Category 1.00. The student's achievement in Category 1.00 is thus not predictable from his profile over Categories 2.00, 3.00, and 4.00.
14. The present study could not explain the somewhat anomalous properties of Category 4.00; why coherent patterns exist only over Categories 2.00, 3.00, and 4.00; why normal achievers in any classification fail to yield coherent patterns; and whether the lack of stronger relationships could be attributed to insensitivity of the criterion measure or to a measure of ambiguity inherent in the structure of the higher levels of the Taxonomy.

Strengths of the Study

The present study was conducted with the participation of many secondary schools throughout the province of Ontario, and thus avoids the shortcomings of small local studies. Since a random sample of schools was used to obtain data, considerable confidence may be placed in the generality of the results.

The design of the study permitted adequate control for ability, with the terms under- and overachiever defined rigorously.

The criterion instrument, upon which the findings depend, has been constructed according to the _Taxonomy of_ Educational Objectives, Cognitive Domain, a model which has been the subject of substantial research which has been reviewed comprehensively in the study. The criterion instrument has been subject to three major revisions prior to use; the trustworthiness of the findings are therefore enhanced by the use of a rigorously developed instrument.

The present study has contributed to the empirical validation of the _Taxonomy_ and the Inventory of Choices. Some findings which have not been reported previously are disclosed. The investigation has also brought to light problems and areas of study requiring further research.

Limitations of the Study

The findings are limited to some extent by the apparent insensitivity of the criterion instrument, which was somewhat difficult for the sample studied. The difficulty may be attributed to the fact that the items were pretested on groups which were not representative of the students who wrote the final edition of the test. Insensitivity is also attributable to the relatively small number of items comprising some subtests. Items in the higher categories of the _Taxonomy_ require considerable deliberation, and thus the number of items that can be answered in the customary testing time allotted to achievement tests is necessarily small. The insensitivity of the criterion measure may be attributable in part to the nature of the _Taxonomy_ itself.

The findings are also limited by the fact that a thorough cross-validation of the criterion instrument was not carried out.

It must not be forgotten that low reliability and a relatively large standard error of measurement were observed for the Category 4.00 subtest. Conservative interpretation of results involving Category 4.00 in this study is indicated.

Educational Implications of the Study

The findings of the study have some significance to the practice of education in the province of Ontario. It appears that the problem of excessive concentration on factual material in the teaching of chemistry may not be so serious as generally supposed; substantial achievement in Comprehension and Application has been noted. There is some question, however, whether competence in the Analysis area is being taught at the Grade 12 level, or, if taught, is being mastered by students. The question also arises as to whether competence in this area presupposes a level of maturity and experience not yet attained by the students; one might also debate whether such high level cognitive ability can be expected of more than a small portion of the population.

The abilities measured by the criterion instrument are not highly correlated with scholastic aptitude or with student grades in chemistry, and seem not to be influenced by such factors as sex of student, choice of textbook, language spoken at home, and the majority of teacher characteristics commonly thought to influence the assessment of student proficiency. For these reasons, tests such as this criterion instrument can be valuable as independent assessments of a student's achievement and as devices for revealing individual student strengths and weaknesses in the attainment of various cognitive objectives which are widely accepted as valuable for science students.

There is also some indication that certain types of students achieve at a relatively higher level in Application than do other students, and perhaps at the expense of achievement in Analysis. The patterns studied in the present investigation also reveal that attainment of chemistry knowledge is not tied to achievement of higher level competencies; conversely, achievement in the higher level cognitive objectives does not seem to be dependent on the quantity of factual material mastered by the student.

The importance of the chemistry teacher, the school, or a combination of both has been emphasized by the study. While no definite characteristic could be singled out, the school environment's contribution to the explainable variance in chemistry achievement and the variable's interaction with mathematics aptitude and verbal aptitude is worth noting.

The present research is more suggestive than definitive; for this reason the educational implications can be regarded as only tentative pending further research.

Suggestions for Further Research

A number of findings in the present study were left unexplained. Further research is needed to explain why overachievers and underachievers have congruent patterns of achievement and normal achievers do not; why these congruent patterns are restricted to Categories 2.00, 3.00, and 4.00; and whether the homogeneity of patterns over Categories 2.00, 3.00, and 4.00 is a phenomenon specific to chemistry achievement or one shared by many subject matter areas. A study of patterns of achievement with respect to ability levels of under- and overachievers might shed some light on these problems.

The school environment, as a variable, contributed more to the explainable variance of chemistry achievement than any factor other than mathematics or verbal aptitude, and yet the characteristic or characteristics of the school environment variable could not be found among the many studied. Research could properly be initiated to determine which teacher and school characteristics are effective in explaining chemistry achievement variance; some factors not studied in this investigation and which might provide suitable starting points for further research are teacher personality variables, recency of training, and updating of science background, as well as school socioenvironmental factors.

A relatively broad study such as the present one should be followed by studies which are more restricted in scope and capable of deeper penetration through the use of more sophisticated and rigorous multivariate techniques of pattern analysis.

Extensions of the present study also suggest themselves: the study of patterns of achievement at other grade levels and in other subject matter areas; the addition of Categories 5.00 and 6.00 to the cognitive profiles; longitudinal studies investigating the stability of achievement patterns. Of particular interest would be the study of achievement pattern changes in the transition from high school to university; such a study could shed light on the role of the Grade 13 year in Ontario. A search for patterns in the Affective Domain of the Taxonomy should be attempted.

The recent introduction of a new course of study in Grade 12 Chemistry in Ontario suggests a replication of the present study, with the criterion instrument being modified to embrace the new subject matter; the resulting comparisons could prove most interesting.

Replications could be used to determine whether results of the kind reported in this study apply to a continuing population of Grade 12 Chemistry students. The usefulness of such results is enhanced when some indication of stability is available.

One aspect not attempted in the present research was the examination of patterns of achievement among low, average, and high achievers, as distinct from the patterns of under- and overachievers; little previous study has been carried out in this area.

Further validation of the Taxonomy is needed. The dependence upon factual knowledge of items in higher categories confounds attempts to measure achievement at these higher levels; attempts to hold knowledge constant by presenting a reading passage have not met with unqualified success, nor has the attempt to equalize difficulty over Taxonomy level been highly successful as an alternative approach to this problem. The somewhat anomalous properties of the Analysis level, the variation in competence attained at the Knowledge level, and the characteristics of patterns revealed in the present study suggest category groupings such as 1.00 versus not 1.00, not 4.00 versus 4.00, and 1.00, 2.00 and 3.00, and 4.00 as alternatives to examining individual categories in further studies.

The decreasing correlation of IQ with increasing Taxonomy level observed by both McFall and Thomas, and the low correlation of scholastic aptitude with Analysis observed by the present writer suggest that tests of intelligence or scholastic aptitude may not be emphasizing adequately the higher cognitive abilities. Closely connected to this problem is the anomalous nature of the Category 4.00 subtest observed in this study. The nature of the tasks comprising Analysis and other higher level competencies need further study; the relationships of Taxonomy subtest scores to measures of reasoning and critical thinking definitely should be investigated.

There is evidence that Taxonomy-type tests must be built to much more stringent specifications than are prevailing achievement tests if they are to be sufficiently sensitive for use as differential tests of educational achievement. Since research findings in this area ultimately rest on the quality of the criterion instruments, great care should be taken in their development. Such care would normally require the use of items selected from a large item pool, and pretested on a sample representative of the target population. The problem of item difficulty would have to be managed carefully, and quite likely the _Taxonomy_ category subtests would have to be of more than customary length. The use of test committees conversant with _Taxonomy_ research findings to assemble such tests would minimize idiosyncratic interpretation of Taxonomic levels in the assignment of items.

The problem of controlling dependence upon knowledge in the higher levels of the _Taxonomy_ may be mitigated by the use of tests consisting of sections which are sealed after use and before proceeding to sections requiring responses at higher taxonomic levels. Recent advances in the use of branched tests and on-line computer testing appear to provide fruitful approaches to investigating the further research suggested in the present report.

APPENDIX A

THE ONTARIO GRADE 12 COURSE OF STUDY IN CHEMISTRY

The course of study in chemistry which was in force at the time the data for this study were collected is reproduced on the following pages. The chemistry course of study was bound in one publication with the courses of study in Grade 11 Physics, and Grades 11 and 12 Agricultural Science. These other courses are not reproduced here.

A new course of study in chemistry (Curriculum S.17D) was introduced in September, 1967.

Curriculum S. 17
3,500 — July, 1954 — 1631

ONTARIO
Department of Education

Courses of Study

Grades XI and XII

SCIENCE
AND
AGRICULTURAL SCIENCE

Reprinted without change from Curriculum S. 17,
printed in July, 1952.

Issued by Authority of
The Minister of Education

COURSES OF STUDY
for
GRADES XI AND XII
in
Collegiate Institutes, High and Continuation Schools

SCIENCE

The Science of Grades XI and XII is an experimental study, and emphasis should be based on pupil experiments throughout the course. Accuracy and precision in making observations, taking measurements, and reaching conclusions are the main desiderata. Encouragement should be given to the recording of experiments by means of simple line diagrams, supplemented by very brief notes. Time should not be wasted in writing notes from dictation or in copying material from text or manual.

PHYSICS

OBLIGATORY COURSE

CHEMISTRY

Changes of state (7 periods)

The three states of matter and their general characteristics. Melting and freezing, illustrated by water and naphthalene. Melting points as characteristic physical constants. Evaporation and condensation, illustrated by water or carbon tetrachloride.

Boiling points as characteristic physical constants; influence of barometric pressure on boiling point.
Sublimation, illustrated by iodine, benzoic acid or naphthalene. Effect of temperature on rate of evaporation of water. Effect of humidity.

Changes in volume and energy accompanying changes of state. Explanation of changes of state in terms of the molecular theory of matter.

(It should be stated that individually distinct molecular particles are not thought to exist under ordinary conditions for such substances as metals, salt, diamond, etc.).

The use of characteristic physical properties (density, melting point, boiling point, ability to form solutions, etc.) for identification of substances.

Mechanical mixtures. (6 periods)

Study of such mechanical mixtures as iron and sulphur; copper filings and charcoal; clay and water; kerosene and water; sugar and sand. This should include a discussion of (i) properties of mixtures in relation to properties of the constituents, (ii) methods of separation.
Study of naturally occurring mixtures.. e.g. lake-shore sand, milk, tomato juice.
Methods of separation of mixtures industrially; e.g. settling, filtering, centrifuging, froth flotation, magnetic separation, distillation.

Elements and compounds; Simple chemical reactions. (6 periods)

The distinction between physical and chemical change; a chemical change may be simply described as a process in which one or more new substances are produced.
Study of (i) heating of mercuric oxide, (ii) heating of bluestone, (iii) electrolysis of water, as simple chemical changes.
Law of conservation of mass applied to chemical changes. Simple experiments.
Law of definite proportions. This should be illustrated by such experiments as (i) decomposition of mercuric oxide, (ii) combination of magnesium and oxygen, (iii) decomposition of bluestone to anhydrous copper sulphate and water vapour.
The most important characteristic of a chemical substance is that it has a fixed composition.
Elements and compounds. The experimental criterion of an element is that it is not composed of two or more other substances. Compounds are made from elements combined in definite proportions by weight. The properties of compounds are likely to differ from those of the constituent elements.

Oxygen. (8 periods)

Occurrence of the most abundant element in the free state and in compounds.

Laboratory preparation of oxygen by heating a mixture of potassium chlorate and manganese dioxide. Catalytic action of the manganese dioxide.

Industrial production by distillation of liquid air. Demonstration of the approximate percentage of oxygen by volume in air. Physical properties of oxygen.

The combustion in oxygen of charcoal, sulphur, phosphorus, magnesium, sodium, and iron. Properties of oxides of these, (state, colour, solubility in water, effect of solutions on litmus). This will require brief mention of acids and bases.

The combustion of compounds, for example, kerosene or alcohol, pyrite or sugar, showing the products formed. The combustion of foods.

Importance and uses of oxygen.

Meaning of terms combustion, exothermic, endothermic, kindling temperature, low-temperature oxidation, spontaneous combustion, heat of combustion, catalyst, oxidation.

Air and the production of nitrogen.
(4 periods)

The importance of air, its composition, (nitrogen, oxygen, rare gases, water vapour, carbon dioxide, dust particles). Processes tending to regulate the amount of carbon dioxide in the air.

The carbon cycle. Interdependence of plants and animals. Production of nitrogen from air. Physical properties, importance, and uses of nitrogen.

Reacting weights and atomic weights.
(8 periods)

The reacting weight of a substance (element or compound) is the number of parts by weight of that substance which reacts with 16 parts by weight of oxygen or with the reacting weight of some other substance. A substance may have several reacting weights; such weights are in the ratio of simple whole numbers. (Equivalent weights are defined in exactly the same terms as reacting weights but with reference to 8 parts by weight of oxygen). It should be stressed that the choice of 16 or 8 for oxygen is arbitrary.

Law of Reacting Weights, — the weights of substances (elements or compounds) which take part in a chemical reaction are in the ratio of their reacting weights or multiples of them.

The atomic theory of John Dalton as an explanation of this law. The atomic weight of an element is a selected reacting weight and is based on the atomic weight of oxygen taken as 16. Atomic weights of the common elements may be introduced at this time.

Symbols, formulae and equations.
(10 periods)

The use of the symbol to denote the element and also to represent one atomic weight of the element.

The use of the formula to indicate the elements and their proportions in a compound, and also to indicate the molecular weight, where known, of the compound. For substances whose molecular weights have not been determined, the term formula weight is preferable.

Nomenclature of binary compounds.

Valence, — an indication of the combining power of the atom of an element. The use of the chemical bond as a convenient method of illustrating valence. Simple structural formula for hydrogen, chlorine, oxygen, nitrogen, hydrogen chloride, water, ammonia, methane, and carbon dioxide.

Chemical equations for simple reactions considered thus far.

Simple problems (i) to determine percentage composition from formulae, (ii) to determine formulae from percentage composition, (iii) to determine weights of reactants or products involved in these simple reactions.

Hydrogen.
(7 periods)

Preparation (i) by electrolysis of water, (ii) by action of water or steam on metals, (iii) by reaction of zinc and dilute sulphuric acid. Physical properties of hydrogen. Burning of hydrogen in air and explosion with oxygen.

Uses of hydrogen.

Demonstration of the reaction between hydrogen and hot cupric oxide to illustrate the law of definite proportions and to find the composition of water (method of Dumas).

Water.
(6 periods)

Occurrence and distribution.

Natural water and preparation of pure water.

Properties of chemically pure water (density, boiling point, freezing point, etc.).

Production of a potable water supply.

Dehydration of copper sulphate pentahydrate (bluestone), and sodium carbonate decahydrate (washing soda). Water of hydration. Efflorescence. Anhydrous copper sulphate as test for the presence of water.

Hygroscopic materials; silica gel, concentrated sulphuric acid, glycerin, calcium chloride. Deliquescence of solids.

Solutions.
(6 periods)

Comparison of characteristics of solutions with those of mechanical mixtures.

Examples of solutions. These should be varied enough to show the existence of solutions in different physical states: e.g. air, including water vapour; low-melting alloys; gold and copper alloys; oil or grease in carbon tetrachloride; DDT in kerosene; carbonated beverages.

Meaning of terms: solvent, solute, solubility (relative and quantitative); saturated, unsaturated, and super-saturated solutions; solubility curve (to illustrate change of solubility with temperature). Factors which affect the rate of solution.

Acids and bases.
(3 periods)

Review the effect of acids on litmus.

Further properties of acids (dilute), (1) effect on indicators (2) action on carbonates, (3) action on suitable metals (magnesium), (4) taste as shown by soda-water, vinegar, sour milk, etc.

The effect of bases on the same indicators which were used for acids. The action upon litmus of the solutions of the oxides of the substances already burned in oxygen, and classification as acidic or basic oxides.

Testing a number of substances found in the household to classify them as having acidic or basic or neutral properties.

The reaction of an acid with a base to form a salt and water (neutralization).

Nomenclature of some oxy-acids and their salts, e.g. sulphates. sulphites, nitrates, carbonates, phosphates, chlorates, etc.; ammonium and hydroxide radicals.

Chemical equations and simple problems.

Determination of molecular weights. (12 periods)

The barometer and measurement of atmospheric pressure. Units of pressure: mm. or inches of mercury, atmospheres (one standard atmosphere = 760 mm. of mercury).

Boyle's law, experimentally demonstrated.

Charles' law, experimentally demonstrated. The Absolute Temperature scale.

Problems involving the above gas laws. Use of Standard Temperature and Pressure, (S.T.P.).

Reacting Volumes of gases, e.g. hydrogen and oxygen, demonstrated by the eudiometer. Gay-Lussac's law of combining gas volumes. Avogadro's principle as an explanation of the law of combining gas volumes, and as a proof of the existence of certain diatomic gas molecules, e.g. hydrogen and oxygen.

The diatomic oxygen molecule fixes the molecular weight of oxygen at 32. The volume of 32 grams of oxygen at S.T.P. is $32/1.429 = 22.4$ litres. By virtue of Avogadro's principle this volume of any other gas must contain the same number of molecules, and therefore a molecular weight of that gas. This is the experimental method of finding molecular weights for many substances. Use of the terms gram-molecular volume, gram-molecular weight or mole.

The use of molecular formulae for gases and vaporizable substances, and the information conveyed.

Problems involving (i) calculation of molecular weights with the aid of the gas laws, (ii) calculation of volumes of gases produced in chemical reactions.

Determination of Atomic weights. (4 periods)

Atomic weights are not obtained directly by experiment, but are chosen as the correct fraction or multiple of a reacting weight to correspond to an approximate atomic weight found (i) by application of the Dulong and Petit rule for specific heats, or (ii) Cannizzaro's method, which was to select the smallest weight of the element found in a gram-molecular weight of any compound of that element.

Note the relationship: $$\frac{\text{Atomic weight}}{\text{Equivalent weight}} = \text{Valence}$$

Carbon and its compounds. (7 periods)

Sources and properties of the different forms of carbon. Allotropism.

Uses of carbon in its various forms for lubrication, fuel, reduction, adsorption, etc.

The properties and uses of carbon dioxide reviewed.

The preparation of carbon dioxide by the action of acids on carbonates and a detailed study of its properties.

The action of baking soda in a baking powder. The effect of pressure on the solubility of carbon dioxide in water (Henry's Law).

The action of heat on carbonates.

The sources of carbon monoxide; dangerous and useful properties. The preparation, properties and uses of acetylene. Presence of carbon in fats, carbohydrates, and proteins.

Fuels.
(6 periods)

General survey of solid, liquid, and gaseous fuels.
Heat of combustion — a transformation of chemical potential energy to heat energy.
A comparison of the calorific value of various fuels.
Atomic Energy — comparison with molecular energy.
Discussion of its potentialities and Canada's position as a supplier of fissionable material. The destructive distillation of coal, reference to the important products obtained. A demonstration of fractional distillation; reference to its application in the refining of petroleum.

Sulphur and its compounds.
(8 periods)

Sources of sulphur.
The preparation of the allotropes (rhombic, monoclinic, plastic).
Properties and uses of sulphur.
Demonstration of the preparation of hydrogen sulphide and its use in the preparation of metallic sulphides.
(Note the tendency of some of these sulphides, such as arsenic, antimony and zinc, to pass through filter paper.)
The laboratory preparation of sulphur dioxide. The properties of its solution and its uses, e.g. bleaching and the production of sulphites (chemical wood pulp).
The principles of the commercial production of sulphuric acid.
The properties and uses of sulphuric acid.
References to such sulphates as those of calcium, copper, magnesium, and sodium.

Common salt.
(4 periods)

A brief discussion of the commercial recovery and industrial importance of salt.
A study of its properties.
A study of the reaction of sulphuric acid with salt.
The laboratory preparation and properties of hydrogen chloride and of hydrochloric acid.

Sodium and potassium.
(3 periods)

The action of air on sodium and on potassium. A review of the reaction of these metals with water.
A discussion of the properties of metals as illustrated by sodium and potassium.
A comparison of the properties of sodium hydroxide and potassium hydroxide.
The flame test for the presence of sodium and potassium.

Halogens.
(9 periods)

A discussion of the production of chlorine by the electrolysis of salt.
Experiments to prepare chlorine in test-tubes by the oxidation of hydrogen chloride (as hydrochloric acid).
A demonstration of the preparation and collection of chlorine and a detailed study of its properties.
An experimental study of the properties of an aqueous solution of chlorine.

A demonstration of the preparation and collection of bromine and an experimental study of its properties.

A demonstration of the relative activity of chlorine and of bromine vapour by comparison of the reactions with antimony, moist blue litmus paper, and solutions of sodium chloride, sodium bromide and sodium iodide.

Commercial sources and uses of bromine.

A demonstration of the preparation and collection of iodine.

A comparison of the properties of chlorine, bromine, and iodine. Reference to fluorine — its importance in dental health.

Compounds of nitrogen.
(7 periods)

The properties of nitrogen.

Laboratory preparation of nitric acid; its acid properties when diluted; its oxidizing action when concentrated; its uses; its toxic effect.

The properties and uses of such nitrates as those of sodium, potassium, ammonium, and calcium.

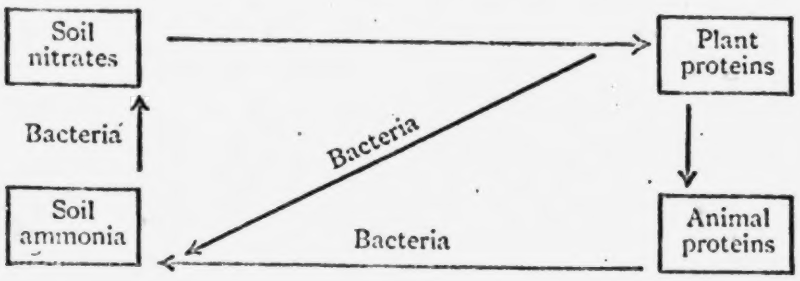
The brown-ring test for nitrates.

Laboratory preparation of ammonia; its properties and uses.

Properties of a solution of ammonia.

Brief discussion of the formation and properties of such ammonium salts as ammonium chloride and ammonium sulphate.

Nitrogen and soil fertility — simple explanation of the nitrogen cycle.



Inert gases.
(2 periods)

The presence of rare gases in the air.

Discuss their chemical inactivity and commercial uses.

Commercial source of helium. Briefly discuss the history of the discovery of these gases, emphasizing the importance of precise and painstaking research.

Calcium and magnesium.
(6 periods)

The reaction of calcium with water.

Occurrence of calcium carbonate ((limestone and marble).

Heating of calcium carbonate. The commercial preparation of quicklime. The slaking of quicklime.

Commercial uses of limestone, quicklime, slaked lime, gypsum, bleaching powder, calcium chloride.

Occurrence of magnesium in dolomite. Reference to Canadian production in Renfrew county (Pidgeon Process).

Properties of magnesium.

It should be emphasized that the metal is resistant to oxidation at ordinary temperatures; only the ribbon or wire forms of magnesium are readily ignited. Importance of magnesium in making low-density alloys such as magnalium.

Rate of reaction. (3 periods)

Throughout the course the attention of students should be directed to instances of the following factors influencing the rates of reactions:
(1) heat (2) light (3) concentration (4) surface area (5) catalysis. At the conclusion of the course a recapitulation of this topic should be made.

Industrial chemistry.

A class which is situated near an industry which uses chemical processes should make a study of those processes whenever practicable, in order to make students realize that chemical reactions are the basis of many of our industries.

APPENDIX B

THE INVENTORY OF CHOICES QUESTION BOOKLET AND ANSWER SHEET

The Inventory of Choices question booklet and answer sheet used in this study differ from the original in that the booklet and answer sheet were printed as separate documents. The method of indicating choices in the answer sheet was arranged to facilitate keypunching.

INVENTORY OF CHOICES

By T.Bentley Edwards & Alan B.Wilson

You are being requested to co-operate in a research study being conducted at the University of Toronto into the relations between students' attitudes toward a variety of topics and their school, recreational and occupational interests. The information you provide will be held in strict confidence.

You are asked to give your opinions about many statements. There are no "right" or "wrong" answers. If you have difficulty deciding on some items, mark the answer which seems closest to what you believe even though you may have doubts. It is important to mark every item.

There is no time limit, but you are expected to finish in a class period, and so do not spend too long on any one item.

How to indicate your choices:

Consider the following sample item:

83. I would rather go sailing than play golf.

This student agrees slightly with statement No. 83 and has indicated his choice by marking an "X" in the proper box under the number 83.

An ordinary pen or pencil may be used to mark your choices.

	83	84
Strongly agree		
Moderately agree		
Slightly agree	X	
Slightly disagree		
Moderately disagree		
Strongly disagree		

Please mark your answers on the separate answer sheet provided. When you have marked all your choices, turn the sheet over and complete the other side.

NOTE THAT QUESTIONS 1 - 5 ARE OMITTED IN THIS EDITION

6. I am more interested in finding out how TV has affected people's taste than in finding out how TV works.

7. If I were employed by a company manufacturing chemicals, I would rather stay in research than become a company executive, so long as the loss in pay was not too great.

8. I find paintings interesting when I am able to see how they represent the artists' attitudes toward life.

9. The opinion of friends helps more than reading in making up my mind.

10. I enjoy swimming in the ocean by myself, or, for safety, with a companion or a life guard, more than swimming in a pool.

11. School mathematics courses should concentrate more on practical consumer and business training.

12. Fabulous IBM machines are used to calculate insurance rates. However, data must be fed into the machines. The interviewing techniques for collecting data interest me more than an explanation of how the machines work.

13. Medical experiments using live animals are cruel and inhuman.

14. I never worry about how things are going to work out -- they usually seem to take care of themselves.

15. When I'm watching a movie I sometimes lose track of the plot because I'm wondering how the lighting and stage effects are worked.

16. I like to ride alone. The feel of a good horse under me, his strength and his rhythm, more than make up for the lack of fellow riders.

17. I would rather teach science than do research.

18. "Hotrod" racing would be fun if you didn't have to know about and work on motors.

19. I should rather be elected to the Student Council than be selected as an honour student in science.

20. I think I should enjoy Longfellow's poem "Evangeline" more if it were told as a love story in modern prose.

21. In our complex industrial civilization a young person should specialize early and stick to it.

22. I prefer chess to checkers.

23. Sailing on a boat would be fun with a group of people, but I don't think I'd care much for it by myself.

24. When talking with my friends in the evening I'd rather talk about people we know and have fun with than talk about religion or philosophy.

25. To my way of thinking, the need to keep a city beautiful to look at is the most important argument in favour of smog control.

26. The forces of nature are subjects for wonder and awe -- not analysis.

27. I usually like to do math problems alone rather than discuss them with others.

28. In a history course I would rather have the reasons why the U.S. didn't join the League of Nations explained to me than try to figure it out.

29. I would rather go sailing by myself than watch a football game.

30. It's a sloppy sailor who'll let his sails flap while he basks in the sun and breathes the crisp salty wetness of the air; the keen sailor watches the wind, studies the tides, and understand details of the rigging.

31. The foreign policy of our government should be based on high moral principles even though this may entail a loss of strategic power or prestige.

32. Visiting a foreign country I would want to see the pageantry and architecture so I would not be interested in knowing in advance about their customs and history.

33. Scientists destroy much of the beauty of nature when they explain away its mysteries.

34. Art should be appreciated intuitively. Analysis destroys its beauty.

35. I would rather study algebra than history even though algebra seems to be almost totally unrelated to any other subject.

36. When I'm studying math or science it is refreshing to take frequent breaks watching TV or talking with a friend.

37. I should prefer to be a machinist rather than a salesman.

38. I never wonder how the time is going when I'm painting a room or sawing firewood like I do when I'm studying math or physics.

39. If pay, housing, etc., were equal, I should like the work of a forest ranger better than that of a minister.

40. I am more interested in following newspaper reports on the recent discoveries regarding "negative matter" than on the developments in racial integration in the schools.

41. I should prefer the live theatre to movies if they were the same price.

42. I spend more of my free time on hobbies like stamp collecting, woodwork, etc., than going to parties or entertaining friends.

43. When I was little I liked erector or meccano sets more than tops.

44. I frequently think about the reasons for other peoples's misbehaviour instead of reacting with irritation.

45. I should rather read and be able to understand William Shakespeare's Hamlet than Michael Faraday's Experimental Researches in Electricity.

46. I can visualize myself reading a paper to a scientific society meeting but not chatting socially in the corridor while a meeting is in progress.

47. Instead of developing expensive tastes, what I would like most to get from my education is either a purpose for my life or an affirmation of my present purposes.

48. Abilities of sign writers are different from the abilities of men who run a sign business. If I had the ability to do either, I would rather learn to run the business than paint the signs.

49. I like history and civics much better than science and mathematics.

50. Science has definitely not been able to show that coloured races are inferior to white races.

51. I don't like being interrupted while I'm doing laboratory experiments by friends who feel like talking.

52. A person should throw himself into life with vitality -- the scientist's reflection on how things work is a wet blanket on the spontaneous pleasures of affection.

53. If I were a musician the thing I should like best about it would be getting across to the audience the basic idea of the composer.

54. In studying about the building of the pyramids, I should be more interested in the engineering feat than in the class structure and economy of Egypt which made such magnificent display possible.

55. I like Dixieland jazz better than "rock and roll".

56. A businessman should make his decisions strictly according to the interests of his business. He should not worry about what happens nationally to wages and prices.

57. Because they need to get close to life, artists are entitled to special consideration if they treat lightly the ties of marriage.

58. I would sooner have a big living room for parties than have a workroom for hobbies.

59. When I see an article about "electronic brains" I am more interested in finding out how they work than what their uses are.

60. When I am buying clothes I pay less attention to the ones that don't show than to the ones that do show.

61. When you go on an automobile trip it is much more fun to pick places to stay as you go along rather than writing ahead for reservations.

62. I should rather take a shop course than a world history course.

63. I would rather be known as the writer of a social column published in many papers than as the Director of an astronomical observatory.

64. If I had an hour to wait for a train I should more likely read The Scientific American than The Atlantic Monthly.

65. An impulsive person is warm and sincere; one who analyzes his emotions is cold and "phony".

66. Chemistry experiments are fun to watch so long as there are plenty of explosions and colour changes.

67. I prefer a science class that is run along fairly formal lines so that I can avoid the distractions arising from personal entanglements with other members of the class.

68. A girl should wear sweaters and "pearls" or whatever most of the girls are wearing rather than conspicuous hand-made jewelry.

69. Nobody should be allowed to cut down the redwood forests and turn them into lumber.

70. Sometimes when a fellow is out with the gang, he pretty well has to do a few things he knows he really shouldn't.

71. If I were interested in studying flowers, I should be attracted chiefly by the beauty of the flowers. Comparing the structure of different kinds of flowers would not interest me much.

72. If I had the necessary athletic prowess I should prefer to excel in the cross-country marathon than in football.

73. In science fiction stories, I like the ones with interesting scientific theories that hang together, even if they're not completely true, better than those about the social problems of space settlements.

74. I enjoy working hard at a science project even if others don't recognize my accomplishment.

75. I like to watch a big house afire.

76. Mercy killing should be legalized for cases of extreme suffering where there is no hope for cure.

77. When I look at the stars at night I sometimes meditate on the way the universe works.

Items 78 - 80 omitted.

Date ____________________ Name ______________________________

(Please Print)

INVENTORY OF CHOICES - ANSWER SHEET

Instructions: For each statement in the accompanying booklet place an "X" in the appropriate box to indicate the answer that corresponds most closely with your attitude toward that statement.

	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Strongly agree															
Moderately agree															
Slightly agree															
Slightly disagree															
Moderately disagree															
Strongly disagree															

	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
Strongly agree																				
Moderately agree																				
Slightly agree																				
Slightly disagree																				
Moderately disagree																				
Strongly disagree																				

	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60
Strongly agree																				
Moderately agree																				
Slightly agree																				
Slightly disagree																				
Moderately disagree																				
Strongly disagree																				

	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80
Strongly agree																				
Moderately agree																				
Slightly agree																				
Slightly disagree																				
Moderately disagree																				
Strongly disagree																				

(Please complete other side)

PLEASE MAKE SURE THAT EVERY ITEM HAS BEEN ANSWERED

Please do not write in these spaces.

P - T																
P - I																
P - A																
T - I																
T - A																
A - I																

PERSONAL INFORMATION

All information entered on this sheet will be kept strictly confidential. Please complete all items.

1.- 5. Name______________________________________
(Please print; surname first)

6. Sex: 1.______Male 7. Age ______years
2.______Female (to nearest birthday)

School______________________________________

Home Form ________________

How many brothers and sisters do you have in your family ? (Write the numbers in the blanks)

8. ______ older brothers 9. _______ younger brothers
10. ______ older sisters 11._______ younger sisters.

12. The chief language spoken in our home is
1. ______ English
2. ______ French
3. ______ Other.

13. My family has lived in Ontario for
1. ______ less than one year
2. ______ less than two years
3. ______ less than five years
4. ______ less than ten years
5. ______ 10 - 24 years
6. ______ 25 - 49 years
7. ______ 50 - 74 years
8. ______ 75 - 99 years
9. ______ over 100 years.

14. My father's (or male guardian's) occupation is
____.

15. My mother's (or female guardian's) occupation is
____.

16. The occupation that I would like to have 10 years from now is
____.

17. The subject I liked best at High School is
1.______ English
2.______ Foreign language
3.______ History and geography
4.______ Mathematics
5.______ Science
6.______ Commercial subjects
7.______ Industrial Arts or Home Economics
8.______ Art
9.______ Music

18. The subject from the above list that I have liked least is number _________.

19. Are you now repeating Grade 12 Chemistry ?
1.______ No
2.______ Yes

20. The way I usually feel about school is
1.______ I like school very much
2.______ I like school somewhat
3.______ I like school slightly
4.______ I dislike school slightly
5.______ I dislike school somewhat
6.______ I dislike school very much.

21. My educational plans are as follows:
1.______ Complete Grade 12
2.______ Complete Grade 13.
3.______ Leave school without completing Gr.12
4.______ Undecided.

AND THEN

22. 1.______ Enter University
2.______ Enter Teachers' College
3.______ Enter a school of nursing
4.______ Enter technical or trade training
5.______ Enter business college
6.______ Obtain a job
7.______ Work at home
8.______ Other plans
9.______ Undecided

Please make sure that both sides of this sheet have been completed. Thank you very much for your co-operation.

APPENDIX C

THE PERSONAL INFORMATION QUESTIONNAIRE

Personal information used in this study was obtained from responses to the questionnaire which forms this Appendix. This questionnaire was printed on the reverse side of the Inventory of Choices Answer Sheet. Students completed both sides of the document at one sitting.

PERSONAL INFORMATION

All information entered on this sheet will be kept strictly confidential. Please complete all items.

1.- 5. Name____
(Please print; surname first)

6. Sex: 1.______Male
2.______Female

7. Age ______years
(to nearest birthday)

School____

Home Form ________________

How many brothers and sisters do you have in your family ? (Write the numbers in the blanks)

8. ______ older brothers 9. _______ younger brothers
10. ______ older sisters 11.________ younger sisters.

12. The chief language spoken in our home is
1. ______ English
2. ______ French
3. ______ Other.

13. My family has lived in Ontario for
1. ______ less than one year
2. ______ less than two years
3. ______ less than five years
4. ______ less than ten years
5. ______ 10 - 24 years
6. ______ 25 - 49 years
7. ______ 50 - 74 years
8. ______ 75 - 99 years
9. ______ over 100 years.

14. My father's (or male guardian's) occupation is
____.

15. My mother's (or female guardian's) occupation is
____.

16. The occupation that I would like to have 10 years from now is
____.

17. The subject I liked best at High School is
1.______ English
2.______ Foreign language
3.______ History and geography
4.______ Mathematics
5.______ Science
6.______ Commercial subjects
7.______ Industrial Arts or Home Economics
8.______ Art
9.______ Music

18. The subject from the above list that I have liked least is number _________.

19. Are you now repeating Grade 12 Chemistry ?
1.______ No
2.______ Yes

20. The way I usually feel about school is
1.______ I like school very much
2.______ I like school somewhat
3.______ I like school slightly
4.______ I dislike school slightly
5.______ I dislike school somewhat
6.______ I dislike school very much.

21. My educational plans are as follows:
1.______ Complete Grade 12
2.______ Complete Grade 13.
3.______ Leave school without completing Gr.12
4.______ Undecided.

AND THEN

22. 1.______ Enter University
2.______ Enter Teachers' College
3.______ Enter a school of nursing
4.______ Enter technical or trade training
5.______ Enter business college
6.______ Obtain a job
7.______ Work at home
8.______ Other plans
9.______ Undecided

Please make sure that both sides of this sheet have been completed. Thank you very much for your co-operation.

Date ____________________________ Name____

(Please Print)

INVENTORY OF CHOICES - ANSWER SHEET

Instructions: For each statement in the accompanying booklet place an "X" in the appropriate box to indicate the answer that corresponds most closely with your attitude toward that statement.

	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Strongly agree															
Moderately agree															
Slightly agree															
Slightly disagree															
Moderately disagree															
Strongly disagree															

	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
Strongly agree																				
Moderately agree																				
Slightly agree																				
Slightly disagree																				
Moderately disagree																				
Strongly disagree																				

	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60
Strongly agree																				
Moderately agree																				
Slightly agree																				
Slightly disagree																				
Moderately disagree																				
Strongly disagree																				

	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80
Strongly agree																				
Moderately agree																				
Slightly agree																				
Slightly disagree																				
Moderately disagree																				
Strongly disagree																				

(Please complete other side)

PLEASE MAKE SURE THAT EVERY ITEM HAS BEEN ANSWERED

Please do not write in these spaces.

P - T															
P - I															
P - A															
T - I															
T - A															
A - I															

APPENDIX D

MATERIALS MAILED TO PARTICIPATING SCHOOLS

This Appendix contains specimens of all materials, other than test booklets and answer cards or sheets, mailed to schools which took part in this study.

Materials are exhibited in chronological order.

1375 Amber Cres.,
Oakville, Ont.,
April 21,1964.

Would you please discuss the following request with your Science Department Head or chemistry teacher:

The Ontario College of Education has just granted approval of my proposed doctoral study which is concerned with patterns of achievement in Grade 12 chemistry. In order to carry out the study it is necessary to administer a chemistry test and a questionnaire, in mid-May, in a number of schools which are representative of the secondary schools of the province. I solicit your co-operation, and in support of this request enclose a letter from Dr. G.E.Flower, Director of Graduate Studies at O.C.E.

The study, which is being conducted with the permission of the Ontario Department of Education, is described briefly in the enclosed information sheets. Please read these fully before making your decision.

I realize that May is a busy time of year and that my request gives you short notice; unfortunately this situation is unavoidable and I ask yourindulgence.

Regardless of the nature of your decision, please return the pink reply sheet as soon as possible. Thank you.

Yours sincerely,

Alexander Even,
Head, Science Dept.,
Oakville-Trafalgar High School.

Ontario College of Education
UNIVERSITY OF TORONTO
371 Bloor Street West, Toronto 5

April 21, 1964

To Headmasters of Selected Ontario Secondary Schools

Gentlemen:

This letter is going forward with a request from Mr. A. Even, Head of the Science Department of the Oakville-Trafalgar High School and a doctoral candidate here at the University of Toronto, that your school participate in a study of Grade XII Chemistry entitled "Patterns of Achievement in Chemistry and Their Relationship to Personal, Attitudinal and Environmental Factors".

We believe that this will turn out to be a most significant study. The Department of Education through its officials in Secondary Education is fully aware of the study and its dimensions, and likewise considers it worthwhile.

We hope that you will find it possible and indeed useful for your school to participate.

Thank you for your interest and cooperation.

Sincerely yours,

George E. Flower,
Director of Graduate Studies

GEF:ca

Grade 12 Chemistry Study

Information Sheets

The Purpose of the Study

The study proposes to examine patterns of achievement in Grade 12 Chemistry and their relationship to personal, attitudinal and environmental factors (the latter including such matters as teaching load, size of school, etc.).

Conditions under which the Study will be Conducted

1. Participation in the study will be at no cost to the co-operating schools.
2. All data will be treated in strict confidence. Names of schools, principals, teachers and students will not be divulged to any source, including the Ontario Department of Education. (the Department of Education concurs in this matter). Names will not appear in any published findings. This is to guarantee that no rating of any school, teacher, or system will be attempted.
3. Schools have been selected at random to give some degree of generality to the findings. There is no other reason for selecting your school.

Data-Gathering Instruments

OTAC - this is a thought-provoking objective-type chemistry test requiring one hour of writing time, plus time for handing out pencils, answer cards, etc. A double period should suffice. The test includes only those topics on the present Grade 12 Course of Study (General Course) to the end of the halogens. This test should be administered in the period from May 11 to May 15 if at all possible. All students in the school should write at one time, to avoid exchange of information. Electrographic pencils, such as those used in the Carnegie Study tests, will be needed.

Answer cards will be machine-scored, interpreted as percentiles and percentage marks, and returned to participating schools during the week of June 1.

Chemistry teachers will be asked to give criticisms of the test, and to complete a short questionnaire.

Student Questionnaire - this includes a survey of student attitudes (the Inventory of Choices) and some personal information similar to that gathered in the Carnegie Study. Administration of this instrument is not timed and should take no more than one class period. A study period could be used, or perhaps the guidance department could donate a period for the purpose. Students need not all answer the questionnaire at the same time.

Follow-up In June, staff members would be asked to enter the students' final chemistry marks and final average marks on prepared lists that will be provided. These marks will be used to check the validity of OTAC.

Experience with the Tryout Form of OTAC

A prototype of OTAC was administered to the students of Oakville's three secondary schools last May, both as a"confidential examination" and to prepare the students for the June examination. Writing this test had a most salutary effect on the pupils, who discovered areas of weakness in good time, and who were able to repair deficiencies before the final examinations began. Our results in June were very much better than we had anticipated.

Our experience is that students must study in advance for OTAC. Should you decide to participate in the study, the students should be informed immediately. You should use what motivating devices you can to ensure that pupils will prepare themselves adequately for the test. Pupils are apt to find the results discouraging if the test is given with insufficient notice.

If you wish to use OTAC as part of an examination at any time in the future, I would be glad to loan you sufficient copies and to provide answer cards free of charge.

I hope that the above description has answered most of your questions. Should you have other questions, whose answers would affect seriously your decision to participate in this study, I will accept collect telephone calls to answer these.

I think that the study will be of mutual benefit, and to this end intend to send a full report to each participating teacher, once the data have been analyzed.

Yours sincerely,

Telephones
Home: 844-3448
School: 845-2875
Area code 416.

Alex Even,
Head, Science Dept.,
Oakville-Trafalgar High School

REPLY SHEET

To A. Even
1375 Amber Cres.,
Oakville, Ont.

1. I wish / do not wish to participate in the study described in your letter of April 21, 1964.

2. The number of academic Grade 12 chemistry students that will participate is ____________.

3. The number of teachers teaching Grade 12 Chemistry in this school is ___.

4. Date when OTAC will be administered (double period required) __________.

5. Date when questionnaire will be administered (single period required) ___________.

6. Person to whom testing materials are to be sent:

   Name ______________________________

   Position ______________________________

7. Person to whom results are to be sent:

   Name______________________________

   Position ______________________________

8. Results of the chemistry test to be received on or before __________1964.

Date ___________.

Signed ______________________________

Position ______________________________

School ______________________________.

Please return this sheet in the stamped, addressed envelope provided. Thank you.

1375 Amber Cres.,
Oakville, Ont.,
May 6, 1964.

Dear Colleague:

Thank you very much for offering to participate in my study. Testing materials and questionnaires are now going forward to you under separate cover, and should reach you in time for administration on the dates requested. The materials are being shipped via

Please note the following:

1. The IBM answer cards for OTAC must be returned promptly. Processing of cards can not be done until all cards are returned.

2. Should the tests arrive a few days' ahead of schedule, please make sure that the students are not "coached". The effectiveness of the results depends on all students having little or no inkling of the test content and tasks required of them. If students have been "primed" the results will be untrustworthy.

3. Please do not retain any test booklets or questionnaire folders, and please do not make copies of them. I will gladly lend you any number of copies should you wish to use these materials at some future date.

4. If the parcels arrive undamaged, and are opened carefully, the wrappings and boxes can be used to return the materials. Enclosed in this letter is a return address sticker and return postage.

5. Two report forms (shipped with the materials) are to be completed and returned with the tests and questionnaires.

Your help is much appreciated.

Yours sincerely,

Alex Even.

Packing List

This shipment contains

_________ OTAC test booklets

_________ IBM answer cards for above

_________ Administration Instructions for OTAC (light yellow)

_________ Questionnaire folders

_________ Answer sheets for questionnaire

_________ Administration Instructions for questionnaire (green)

_________ Return Instructions (pink)

_________ OTAC Testing Record (orange-yellow)

_________ Questionnaire Report Form (blue)

Please check contents immediately and report any discrepancies by telephone (collect).

School - 845-2875
Home - 844-3448
Area Code 416

A. Even.

ONTARIO GRADE 12 CHEMISTRY STUDY

OTAC Administration Instructions

1. Students should receive the following at least five minutes before the test is to begin:

   (a) One IBM answer card

   (b) One electrographic pencil with a serviceable eraser attached.

   (c) One OTAC question booklet which must not be opened until you give the signal to start. Make sure that the students understand this point.

   (d) One piece of scratch paper or foolscap, for rough work.

2. Have the students enter name, etc. on the answer card according to the instructions on the front cover. Note that the blank labelled "test form " is to be used for other information, and that some blanks on the card are not to be filled. The electrographic pencils must not be used for entering this information.

3. Since most Grade 12 students have had considerable experience with Carnegie and Ontario Departmental Tests, further instructions will probably not be needed.

4. Caution the students not to make any marks in the test booklet and give the signal to begin.

5. Please make no attempt to interpret any item for the student.

6. This is a timed test: exactly one hour (with no breaks) should be allowed for answering the items. Make sure that all students stop at the end of one hour, even if many appear not to have completed the test. This is the only way by which comparable results may be obtained.

7. Have all the students stop writing immediately after you give the signal to stop. Please make sure that no question booklets are retained by students. This is very important.

8. After all materials have been collected, please complete the OTAC Testing Record (orange-yellow sheet) and return with the question booklets and answer cards.

Thank you very much for your help.

A.Even.

ONTARIO GRADE 12 CHEMISTRY STUDY

Questionnaire Administration Instructions

The questionnaire consists of two parts:

1. The Inventory of Choices in which students are asked to indicate their agreement or disagreement with 72 statements in the folder. A separate answer sheet is used for responses.
2. A personal information survey which is found on the reverse side of the answer sheet used for the Inventory of Choices.

Administration of the questionnaire is very simple: each student is given an Inventory of Choices folder and an answer sheet, and is then asked to supply the information sought. The Inventory of Choices should be completed first. While no time limit is specified, a normal class period should be adequate for completing the questionnaire. The instructions given on the front cover are self-explanatory.

Please stress the following points to the students:

1. Names are used only for the purpose of collating results. All answers are kept strictly confidential, and no names will be used in any reports.

2. All items should be answered, particularly in the Inventory of Choices. Even though some statements may be difficult for some students to decide upon, a best attempt should be made to answer the items.

3. In the personal information survey, where a short blank occurs an "X" or check mark ( ) should be used to indicate the choice, unless a number is asked for.

Should any serious question concerning the administration of this questionnaire arise, please call me collect at 845-2875 (school) or 844-3448 (home) (Area Code 416).

A report form for this questionnaire (blue sheet) is enclosed. Please complete this form and return it with the folders and answer sheets.

Thank you very much for your co-operation.

A.Even.

Return Instructions

1. Please return all materials as promptly as possible. Parcels should be well-wrapped and securely tied with string to protect the contents. The original wrappings and boxes can be used again in most cases. Return postage and a return address sticker were enclosed in my letter dated May 6, 1964.

2. The IBM answer cards should be wrapped in a sheet of paper before being enclosed in the parcel. Damaged or badly-smeared cards cannot be machine-scored.

3. Please do not retain any copies of OTAC or the questionnaire, and please do not make any copies of them. One copy retained and later accidentally lost could eventually invalidate a whole year's testing programme, should another sample be tested in the future. If you wish to use these materials again, I will gladly send you (on loan) as many copies as you need.

4. The OTAC Testing Record (orange-yellow sheet) and the Questionnaire Report Form (blue sheet) should be completed and returned with their respective booklets.

Many thanks for your help.

A.Even.

OTAC Testing Record

Please complete this form and return it with the OTAC booklets.

1. Date administered

   Time of administration

   Number of students participating

2. Please report in this space any deviation from normal testing procedure that might influence the results of this study:

3. Please use this space to comment on the suitability of OTAC and student reaction to it. Any suggestions or criticisms of this test will be welcome. (Use the reverse side if necessary).

4. Do you wish to use OTAC as part of your final examination in Chemistry in June 1964?

   If your answer is "yes" keep the OTAC test booklets until after the examination, but return all other materials now. Additional answer cards and scoring stencils will be sent to you, free of charge. Indicate here the number of additional IBM cards required.

5. Do you wish to use OTAC as part of an examination or as a test at any other time?

   If your answer is "yes", please specify the probable date and number of students likely involved.

6. Name

   School

Thank you. A. Even

Questionnaire Report Form

Please complete this form and return it with the questionnaire folders and answer sheets.

1. Date administered

   Time of administration

   Number of students responding

2. Please report in this space any deviation from normal procedure, or any other circumstances that might influence the results of this study:

3. If you wish, use this space to comment on the questionnaire and student reaction to it. Any suggestions or criticisms will be welcome.

4. Name

   School

Thank you for your co-operation

A.Even

1375 Amber Cres.,
Oakville, Ont.,
May 1964.

Dear Colleague:

This letter will acknowledge safe return of the materials recently sent to you in connection with my Grade 12 Chemistry study. I will endeavour to have the results of the chemistry test in your hands by June 1.

In the meantime would you please complete, or have completed by each teacher of Grade 12 Chemistry, the questionnaire which is enclosed? I have included a stamped envelope and a copy of the questionnaire for each teacher of the subject. I would appreciate it very much if all copies were returned by June 1 at the latest.

Many thanks for your assistance.

Yours sincerely,

Alex Even.

Ontario Grade 12 Chemistry Study

TEACHER QUESTIONNAIRE

Each teacher of Grade 12 Chemistry will please complete one copy of this questionnaire. All replies will be kept confidential. Please return these sheets, in the stamped envelope provided, by June 1 at the latest.

1 - 5. Name ____

School ____

Position ____

6. Sex ________.

7. OSSTF Category ________ (mark "NA" if not applicable).

Teaching Experience:

8 - 9. Grade 12 Chemistry ________ years (including this year).

10 -11. Grade 13 Chemistry ________ years (including this year).

Teaching Load :

Please list all classes in all subjects taught (e.g. 12 B Chem., 9 C Sc.)

Class	Number of Students in Class	No. of periods per week and length of period.

12.

13. Grade 12 Chemistry textbook used: (please indicate by an "X")
    1. ______ Croal. Couke & Louden: CHEMISTRY FOR SECONDARY SCHOOLS
    2. ______ Cragg, Graham & Young: THE ELEMENTS OF CHEMISTRY
    3. ______ Other (please specify authors and title ________________ ____
    4. ______ None.

14. Laboratory Manual Used in Grade 12 Chemistry:
    1. ______ Croal, Couke & Louden: EXPERIMENTS IN LABORATORY CHEMISTRY
    2. ______ Motherwell & Young: THE ELEMENTS OF CHEMISTRY IN THE LABORATORY
    3. ______ Herron & Maddeford: A FIRST LABORATORY MANUAL IN CHEMISTRY
    4. ______ Other (please specify authors and title)________________ ____
    5. ______ None.

15. Have you used any of the new audio-visual aids or instructional devices this year ? (e.g. CHEM study films, models, programmed learning, etc.) ___________. If yes, please list these on the reverse side of this sheet.

16. What do you consider to be the objectives of the present Grade 12 Chemistry course ?

17. What do you suggest the objectives of the Grade 12 Chemistry course should be ?

18. What would you regard as the objectives, in general, of the Grade 12 Chemistry teachers of Ontario ?

Thank you for your co-operation.

1375 Amber Cres.,
Oakville, Ont.,
May 30, 1964.

Dear Colleague:

Enclosed are the results of the OTAC (chemistry test), listing for each participating student the raw score and percentile rank. Two tables for converting percentile ranks to percentage marks are also enclosed, as well as a note on percentiles and use of these tables.

You are, of course, free to use the results in any manner you wish. Since the raw scores were, in general, lower than anticipated, it may be wise to give the students only their percentile ranks, or the school marks derived from the tables, if these are used.

I shall be contacting you about June 15 to obtain lists of final marks in chemistry and average final marks of those students taking part in the study. In the meantime would you please report any discrepancies found in the students' names on the printed lists ?

Yours sincerely,

Alex Even.

A NOTE ON PERCENTILE RANKS AND THE USE OF THE ATTACHED TABLES

Interpretation of Percentile Ranks:

The percentile rank shows the individual's standing within a specified group by indicating the percentage of those who obtained a lower score. Thus a student with a percentile rank of 90 has obtained a score higher than that obtained by 90 per cent of the group. The group used in this study is believed to be representative of the population of Grade 12 Chemistry students of Ontario.

Use of the Conversion Tables:

Percentile ranks may be converted to school marks by using either of the tables attached. Table I assumes a mean of 62.5 % with a 20 % failure rate and 20 % of the students receiving first class honours. Table II assumes a mean of 65 % with a 15 % failure rate and 20 % of the students receiving first class honours.

Both the above systems of scaling are in common use in Ontario.

TABLE I

CONVERSION OF PERCENTILE RANKS TO SCHOOL MARKS

Assumed mean = 62.5 Failure rate 20 %

Percentile Rank	School Mark	Percentile Rank	School Mark	Percentile Rank	School Mark
99.5-99.9	100	73-75	72	16	48
99.4	99	71-72	71	14-15	47
99.2-99.3	98	69-70	70	13	46
99.0-99.1	97	66-68	69	11-12	45
98.8	96	64-65	68	10	44
98.6	95	61-63	67	9	43
98.4	94	59-60	66	8	42
98- 98.2	93	56-58	65	7	41
97	90	53-55	64	6	40
96	88	50-52	63	5	38
95	87	48-49	62	4	37
94	85	45-47	61	3	35
93	84	42-44	60	1.8-2	32
92	83	40-41	59	1.6	31
91	82	37-39	58	1.4	30
89- 90	81	35-36	57	1.2	29
88	80	32-34	56	.9-1.0	28
87	79	30-31	55	.7- .8	27
85- 86	78	27-29	54	.6	26
83- 84	77	25-26	53	.5	25
82	76	23-24	52	.4	24
79- 81	75	21-22	51	.3	22
78	74	19-20	50	.2	20
76- 77	73	17-18	49	.1	17

Data reproduced from DT: 1961-62: June, 1962, Memorandum 17, with the permission of the Department of Educational Research, Ontario College of Education.

TABLE II

CONVERSION OF PERCENTILE RANKS TO SCHOOL MARKS

Assumed mean = 65 Failure rate 15 %

Percentile Rank	School Mark	Percentile Rank	School Mark
99.8 and over	100	44-46	63
99.6-99.7	99	41-43	62
99.4-99.5	98	38-40	61
99.2-99.3	97	35-37	60
99.0-99.1	96	32-34	59
98.8-98.9	95	29-31	58
98.6-98.7	94	27-28	57
98.4-98.5	93	25-26	56
98.2-98.3	92	23-24	55
98.1	91	21-22	54
98.0	90	19-20	53
97	89	17-18	52
96	88	16	51
95	87	15	50
94	85	13-14	49
93	83	12	48
92	82	10-11	47
91	81	9	46
89-90	80	8	45
87-88	79	7	44
85-86	78	6	43
83-84	77	5	42
81-82	76		41
79-80	75	4	40
77-78	74		39
74-76	73	3	38
71-73	72	2	37
68-70	71	1.8-1.9	36
65-67	70	1.6-1.7	35
62-64	69	1.4-1.5	34
59-61	68	1.2-1.3	33
56-58	67	1.0-1.1	32
53-55	66	.8- .9	31
50-52	65	.6- .7	30
47-49	64	.5	29
		.4	28
		.3	27
		.2	26
		.1	25

Data reproduced from DT;1962-63: June,1963, Memorandum PR - SM, with the permission of the Department of Educational Research, Ontario College of Education.

1375 Amber Cres.,
Oakville, Ont.,
June 19,1964.

Dear Colleague:

To complete the data required for my study, it will be necessary to ask you to enter some marks on the enclosed lists and return these to me in the stamped envelope provided. Instructions are attached to each list. The buff copy is for chemistry marks and the blue copy for information from office records. If it is not feasible to return these lists by June 30, would you be good enough to send me a note stating when these data could be expected ?

I would like to thank you and your fellow staff members most heartily for participating in my study. I realize that this co-operation has made demands on you at a very busy time of year; your interest and support are much appreciated.

Dr. G.E.Flower, Director of Graduate Studies at the Ontario College of Education, expresses his thanks in the enclosed letter.

I expect to have an interim report of my research ready some time this fall, at which time a copy will be sent to you.

Yours sincerely,

Alex Even.

Ontario College of Education
UNIVERSITY OF TORONTO

371 BLOOR STREET WEST. TORONTO 5

June 17, 1964

To Headmasters of Selected Ontario Secondary Schools

Gentlemen:

The purpose of this note is to express our thanks to you and your staff for your cooperation in connection with Mr. A. Even's study of Grade XII Chemistry entitled "Patterns of Achievement in Chemistry and Their Relationship to Personal, Attitudinal and Environmental Factors".

We particularly appreciate your assistance in what we believe will be a most significant study, and particularly so at this very busy time of year.

Thank you for your interest and cooperation.

Sincerely yours,

George E. Flower,
Director of Graduate Studies

GEF:ca

CHEMISTRY MARK LIST

Two columns have been indicated on the attached student list. Opposite each student's name please enter the following marks:

Column A	Final mark in Grade 12 Chemistry. This should be the mark calculated prior to the promotion meeting (i.e. an unadjusted mark) and should incorporate the OTAC score if OTAC scores were used to form part of the students' marks.
Column B	Student's Final Mark less OTAC score: In order to validate OTAC, a special final mark for each student is needed. This mark must not have incorporated in it, in any way, the student's OTAC result. If the OTAC score formed part of this mark, spurious correlation would result; i.e. the correlation between OTAC scores and students' final marks would be higher than it should be. The mark entered in this column need not be expressed as a percentage, as long as the basis is clearly indicated.

Suppose final marks were calculated as follows:

		One student's marks
Final examination	60 %	50
Other examinations	30 %	23
OTAC score	10 %	8
	100 %	81

Student's final mark (column A) 81 (out of 100)

Student's final mark less OTAC (column B) 81-8 = 73 (out of 90)

In the space below please show, in a manner similar to the above example, how your final marks in chemistry were calculated. If different chemistry teachers in your school used different methods of arriving at a final mark, please indicate the different methods and the corresponding teachers' names.

Please return this sheet with the lists. Thank you.

FINAL MARK LIST

Three columns, headed "X", "Y" and "Z" have been indicated on the attached student list.

Opposite each student's name please enter the following information:

Column X	Final mark in Grade 12 Chemistry. This should be the mark entered on the student's report card.
Column Y	Final average mark.
Column Z	The number of subjects taken by the student and used to compute his average; i.e. exclude any subject whose mark is not included in calculating the student's average. English Composition and English Literature should be counted together as one subject unless each is given a mark out of 100. I am assuming that each subject is marked on on the basis of a maximum of 100. If such is not the case please use the space below to show the weight given to each subject, and return this sheet with the lists.

Please add any missing names to the list, and supply the above information for these students as well.

Thank you for your co-operation.

APPENDIX E

THE DEVELOPMENT OF OTAC

APPENDIX E

THE DEVELOPMENT OF OTAC

Preliminary Editions

Three editions of multiple-choice tests were constructed and their items classified by the writer according to the Taxonomy of Educational Objectives. The tests were administered as part of term examinations, and according to the schedule given in Table E-1.

TABLE E-1

ADMINISTRATION SCHEDULE OF PRELIMINARY TESTS

Test	Number of Items	Time[a] (minutes)	Date	Number of Students	Presentation Order of *Taxonomy* Categories
12 E 2	60	30 - 45[b]	20 Jan 61	109	Serial
12 E 2	60	60	24 Jan 62	124	Serial
12 E 3	90	60	6 Feb 63	142	Scrambled
12 E 4	70	60	24 May 63	313	Serial-spiral

[a] The number of items answered per hour depends on the proportion of items in each category of the *Taxonomy*. Generally, low-category items require less answering time than do items in higher categories.

[b] At option of student

Tests 12 E 2 and 12 E 3 consisted of two entirely different sets of items. Item analyses were used to identify items which discriminated poorly, or which were of unsatisfactory difficulty. These two tests were administered in one school only. In administering 12 E 2 the second time, the present writer circulated among the testees in various rooms in order to observe the test-taking behaviors. In 12 E 2 the items of each *Taxonomy* category were placed in blocks in the test booklets, and it was observed that students intuitively sensed to some degree the special nature of a given block of items and frequently skipped over the entire block. For this reason 12 E 3 presented the items not *en bloc*, but rather in scrambled order of *Taxonomy* category.

For both 12 E 2 and 12 E 3 the proctors were provided with simple report sheets to enable the present writer to estimate whether most students were able to complete the tests in the time allotted.

Items of 12 E 2 were selected without special regard for their difficulties. The increasing difficulty with increasing *Taxonomy* level is quite apparent, as shown in Table E-2.

TABLE E-2

ITEM STATISTICS FOR TEST 12 E 2

Taxonomy Category	Number of Items	Median Point-biserial r (subtest)	Proportion of testees getting items right
1.10	20	.28	.74
1.20 - 1.30	14	.23	.52
2.00	8	.24	.41
3.00	10	.26	.38
4.00	8	.32	.24
Total	60		Average .51

A Kruskal-Wallis one-way analysis of variance by ranks (Siegel, 1965, pp.184-193) shows differences in mean subtest difficulty in 12 E 2 to be significant at the .001 level.

In constructing 12 E 4 and OTAC, an attempt was made to make all subtests approximately equal in difficulty, for reasons discussed in Chapter V. The difficulty of each Taxonomy category of 12 E 4 is shown in Table E-3.

TABLE E-3

TOTAL TEST AND SUBTEST DIFFICULTIES
OF TEST 12 E 4

Category	Proportion of testees getting items right
Total Test	.46
1.00 combined.	.43
1.10	.55
1.20 - 1.30.	.32
2.00	.53
3.00	.54
4.00	.35

The apparent difficulty of Category 1.20 - 1.30 may be due to the speeded nature of the test; many students did not reach items near the end of the test. Test 12 E 4 combined 50 of the most discriminating items from 12 E 2 and 12 E 3, plus 20 additional items not previously used. This combination resulted in a test having representative sampling of course content over almost the whole year's program. Items were arranged in Taxonomy categories as shown in Table E-4. Three high schools participated in this tryout.

TABLE E-4

ALLOCATION OF 12 E 4 ITEMS TO SUBTESTS

Subtest	Items	*Taxonomy* Category	Number of Items
1	1 - 14	1.10	14
2	15 - 28	2.00	14
3	29 - 42	3.00	14
4	43 - 56	4.00	14
5	57 - 60	1.20 - 1.30	14

Content Validation of 12 E 4

Correct responses to the items were obtained by pooling the judgments of the author and his colleagues in the Oakville secondary school system. Items were classified according to the *Taxonomy* by the author. The author's colleagues also offered some suggestions as to the rewording of a few items.

The scientific accuracy of the test was appraised by Dr. R. P. Graham, Dean of Science and Professor of Chemistry, McMaster University, and Dr. J. W. Burns, Professor of Chemistry, University of Western Ontario. Very close agreement as to correctness of responses was obtained; in addition, some changes were suggested to bring a few items into agreement with new concepts being taught at the university level.

Student responses to this test were machine-scored and item-analyzed. Table E-5 shows the results of this item analysis.

TABLE E-5

12 E 4 Main Item Statistics

Category	Number of Items	Mean	SD	Skewness	S.E.	Kurtosis	S.E.	Reliability KR-20
TOTAL	70	31.946	9.345	.415	.138	-.278	.275	.843
1.00	28	12.074	4.350	.424	.138	-.364	.275	.730
2.00	14	7.417	2.593	.109	.138	-.597	.275	.563
3.00	14	8.538	2.550	.005	.138	-.516	.275	.570
4.00	14	4.917	2.345	.179	.138	-.407	.275	.504

Results were generally satisfactory, although elimination of some poor items and a reduction in the total number of items would improve the quality of the test. With three exceptions, items correlated more highly with their subtest scores than with the total score. Only three items were below .20 in difficulty (calculated on the basis of the number of students attempting the item, since the test appeared to be partly speeded).

An overall reliability of .84 was obtained using Kuder-Richardson Formula 20. Low subtest reliabilities may be attributed to the following conditions:

1. Subtests of 12 to 14 items may be too short to sample adequately the spectrum of abilities represented by Taxonomy categories or subcategories. Application of the Spearman-Brown prophecy formula (Cronbach, 1960, p.131) gave projected subtest reliabilities which ranged from .82 to .94.
2. The broad categories of the Taxonomy may show considerable overlap under the method of testing for achievement in those categories.
3. Half the Grade 12 students in one school were handicapped by having a succession of "supply" teachers in chemistry for some months prior to the test, because of a serious illness suffered by their regular chemistry teacher.
4. The motivation of the students at the time the test was administered may have ranged considerably. One school's

Grade 12 population had a record of underachievement.

Improvements Incorporated in OTAC

OTAC resulted from a revision of 12 E 4. The following changes were made as a result of the item analysis and suggestions offered from various sources:

1. The number of items was reduced from 70 to 60, to allow nearly all students time to consider all items.
2. Items were scrambled with respect to their *Taxonomy* classification. This rearrangement was made to minimize the observed tendency of students to omit blocks of items, and not to attempt those near the end of the test.
3. Poorly functioning distractors were rewritten.
4. Suggestions of colleagues as to rewording were followed.
5. A few items containing textbook information that Dean Graham and Professor Burns considered doubtful were replaced.

The proportion of items in OTAC allocated to each *Taxonomy* category by the present writer was the same as in 12 E 4.

Following a recommendation by Toops (1960, pp.265-266), a larger proportion of correct answers to items was placed in positions 4 and 5 than in the other three response positions; this redistribution is shown in Table E-6.

TABLE E-6

DISTRIBUTION OF ITEM RESPONSE POSITIONS IN OTAC

Correct Response Position	Number of items in OTAC using this response position
1	11
2	11
3	12
4	13
5	13
Total	60

APPENDIX F

THE ONTARIO TEST OF ACHIEVEMENT IN CHEMISTRY (OTAC)

This appendix contains a specimen of the test booklet used in this study. Choices considered correct are indicated by o . Beside each item is the *Taxonomy* category to which the item was assigned, after consideration by a panel of judges.

OTAC

Ontario Test of Achievement in Chemistry

1964 Edition

INSTRUCTIONS

1. Do not open this booklet until told to do so.

2. (A) On the back of the answer card provided, enter in ink (ball-point or fountain pen) the date, school and signature in the spaces provided. In the blank marked "Test Form" enter instead the name of your chemistry teacher and the class with which you take chemistry; e.g. " Mr. Jones, 12B ".

   (B) On the front of the answer card PRINT your surname and initials in the space provided along the long side of the answer card. Do not fill in the other spaces.

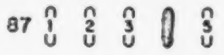
   Do not use the special pencil for entering this information

3. Your answers are to be indicated by marking the answer card with the special pencil provided. Only one choice is to be made per item; more than one choice will cause the item to be counted as incorrect. Erase thoroughly any answer you wish to change.

Example

87. Toronto is a

    1. mountain.
    2. country.
    3. province.
    4. city.
    5. village.



The correct answer is, of course, "city", which is answer number 4. The student has indicated this by making a heavy black mark in the space numbered "4" opposite question 87 on the answer card.

Make no unnecessary marks in or around the answer spaces. Do not rest your pencil on a numbered space while deciding which space to mark. The electric scoring machine cannot distinguish between intended answers and stray pencil marks.

4. You may answer questions even when you are not perfectly sure that your answers are correct, but you should avoid wild guessing, since wrong answers may result in a subtraction from the number of correct answers.

5. Make no marks on this booklet; use the scratch paper provided for your rough work.

DO NOT MARK THIS BOOKLET IN ANY WAY.
DO NOT SPEND TOO MUCH TIME ON ANY ONE ITEM.
ANSWER THE EASIER QUESTIONS FIRST.
FOR EACH QUESTION THE BEST ANSWER IS TO BE CHOSEN.

Part A

1. The process in which a liquid is changed to a gas and back again is
   - 1. condensation.
   - 1.00 2. sublimation.
   - 3. diffusion.
   - 4. evaporation.
   - ○ 5. distillation.

2. Chemical changes are always accompanied by
   - 1. liberation of heat.
   - 2. absorption of heat.
   - ○ 3. energy changes.
   - 4. a gain in weight.
   - 1.00 5. heat and light.

3. Which of the following causes a chemical change ?
   - 1. Defrosting a refrigerator.
   - 1.00 2. Distillation of water.
   - 3. Liquefying air.
   - 4. Adding anti-freeze to a car radiator.
   - ○ 5. Passing an electric current through a solution.

4. Of the following oxides, which gives the most strongly basic solution when dissolved in water ?
   - 1. magnesium oxide.
   - ○ 2. sodium oxide.
   - 3. sulphur dioxide.
   - 4. copper oxide.
   - 2.00 5. zinc oxide.

5. A gas commonly used in electric light bulbs to retard vaporization of the filament is
   - 1. neon.
   - 1.00 2. helium.
   - 3. xenon.
   - ○ 4. argon.
   - 5. krypton.

6. When one buys a pound of Dry Ice, one is really buying
   - 1. water.
   - 1.00 2. hydrogen.
   - 3. nitrogen.
   - 4. oxygen.
   - ○ 5. carbon dioxide.

7. The fractional distillation of liquid air is used commercially to obtain
   - 1. Dry Ice.
   - 2. ammonia.
   - ○ 3. oxygen.
   - 4. carbon dioxide.
   - 1.00 5. hydrogen.

8. All of the following metals readily displace hydrogen from cold water with the exception of
   - ○ 1. magnesium.
   - 2. sodium.
   - 3. calcium.
   - 4. potassium.
   - 1.00 5. two of the above.

9. In a common laboratory method for the preparation of hydrogen from acids
   - 1. no reaction takes place at room temperatur
   - ○ 2. zinc sulphate is left in the generator.
   - 3. the hydrogen evolved is dissolved in water, and this solution is commonly used in subsequent experiments.
   - 1.00 4. copper is added to concentrated sulphuric acid.
   - 5. because of its density hydrogen is collecte by the upward displacement of air.

10. A certain metal will liberate hydrogen from dilute acids, although it does so from water only when the metal is heated strongly and the water is in the form of steam. With relation to the activity series of the metals, this metal
    - 1. is very high in the series.
    - 2. is below mercury and copper.
    - 3.00 3. probably stands close to and below hydrogen.
    - ○ 4. probably stands close to and above hydrogen.
    - 5. is very low in the series.

11. A certain material on extraction with water left a solid residue. The water extract, on evaporation, also left a solid residue. These facts definitely show that the original material was
    - 1. an anhydride.
    - 3.00 2. a hydrate.
    - 3. a compound.
    - ○ 4. a mixture.
    - 5. an emulsion.

12. Which of the following statements is not true ?
    - ○ 1. The solubility of a gas in a liquid usually increases with an increase in temperature.
    - 1.00 2. The solubility of a gas in a liquid usually increases with an increase in pressure.
    - 3. Gases are not very soluble in liquids, although there are many exceptions.
    - 4. Gases always dissolve in one another, if no reaction takes place.
    - 5. Some solid elements can be combined after melting, and on freezing form solid solutions.

13. A substance may be considered to be homogeneous when
    - 1. every particle is the same size as every other particle.
    - 2. the substance is composed of particles whose properties may be quite different.
    - ○ 3. every portion of the substance has the same properties.
    - 2.00 4. the substance is composed of two different elements.
    - 5. all particles of the substance are soluble.

14. Which of the following is not a general property of solutions ?

1. The molecules of the solute are separated from each other as they become distributed throughout the solvent.
2. Solutions are always clear but not necessarily colourless.
3. *1.00* Solutions are always homogeneous after thorough stirring.
4. The homogeneous condition of the solution is permanent.
5. ○ Water is always the solvent.

15. The presence of sugar at the bottom of a cup of cold coffee cannot be explained by the fact that

1. more sugar was put into the coffee than could be dissolved.
2. ○ some sugar precipitated when cream was added.
3. the coffee was not stirred sufficiently.
4. some sugar deposited as the coffee cooled.
5. *3.00* the hot coffee was supersaturated.

16. The amount of evaporation of water into the air of a room may be decreased by

1. increasing the temperature.
2. *1.00* placing the water in a pan of larger diameter.
3. ○ increasing the humidity of a room.
4. decreasing the pressure over the water surface.
5. increasing the molecular motion of the water.

17. Indicate the experimental or observational study giving best support for the statement "Molecules of a gas are in constant motion".

1. Winds exert pressure.
2. *1.00* Heat is necessary to vaporize a substance.
3. Warm air rises.
4. ○ Gases diffuse into one another.
5. Gases are highly compressible.

18. If one volume of oxygen and two volumes of hydrogen are combined at a high temperature, at the same pressure the number of volumes of water vapour obtained is

1. one.
2. ○ two.
3. three.
4. a small fraction of one volume.
5. *3.00* a volume greater than three.

19. Helium has an atomic weight of 4. This means that

1. ○ sixteen atoms of helium weigh as much as four atoms of oxygen.
2. an atom of helium weighs as much as four atoms of oxygen.
3. a molecule of helium contains four atoms.
4. four molecules of helium constitute a gram-molecular weight.
5. *2.00* four atoms of helium constitute a gram-atomic weight.

20. Dalton's Atomic Theory offers a reasonably satisfactory explanation for one of the following statements:

1. When oxygen combines with hydrogen, a large amount of energy is liberated.
2. Forces holding the atoms together in a molecule of nitroglycerine are so small that the compound is explosive.
3. ○ The weight of carbon dioxide produced by burning a given weight of carbon can be predicted accurately if the formula of carbon dioxide is known.
4. *3.00* Hydrogen atoms combine more readily with oxygen atoms than with chlorine atoms.
5. Four atoms of hydrogen combine with one atom of carbon, but two atoms of hydrogen combine with one atom of oxygen.

21. A compound consists of 40 % carbon, 6.7 % hydrogen and 53.3 % oxygen. The simplest formula that the compound could have is

1. ○ CH_2O .
2. C_2HO_3 .
3. C_4HO_5 .
4. CHO
5. *3.00* $C_3H_6O_3$.

(C = 12 H = 1 O = 16)

22\.

$$NO_2 + H_2O \longrightarrow HNO_3 + NO$$

Which of the following sets of numbers, when placed in the same order in the above equation causes it to be correctly balanced ?

1. 1,2,3,4.
2. *3.00* 1,2,2,1.
3. 2,1,2,2.
4. ○ 3,1,2,1.
5. 3,2,1,2.

23. Pentane is C_5H_{12} . When pentane is burned in a plentiful supply of oxygen the products are carbon dioxide and water, and nothing else. How many carbon dioxide (CO_2) molecules are formed from each molecule of pentane ?

1. One.
2. *3.00* Two.
3. Three.
4. Four.
5. ○ Five.

24. The expression

$$4\ FeCO_3 + O_2 \longrightarrow 2\ Fe_2O_3 + CO_2$$

is not considered a balanced chemical equation because it does not conform with

1. the Law of Multiple Proportions.
2. the Law of Definite Proportions.
3. *1.00* Avogadro's Law.
4. Gay-Lussac's Law.
5. ○ the Law of Conservation of Mass.

For Items 25 to 30 refer to the following directions:

DIRECTIONS: Each item consists of an assertion (statement) followed by a reason. In answering select

1. if both assertion and reason are true statements and are related as cause and effect.
2. if both assertion and reason are true, but are not related as cause and effect.
3. if the assertion is true, but the reason is a false statement.
4. if the assertion is false, but the reason is a true statement.
5. if both assertion and reason are false statements.

Directions summarized

	Assertion	Reason	
1.	True	True	Cause and effect relationship
2.	True	True	No cause and effect relationship
3.	True	False	
4.	False	True	
5.	False	False	

4.00 25. Carbon is used in gas masks because carbon is a good reducing agent. O(2)

4.00 26. 24 grams of graphite when burned will yield 88 grams of carbon dioxide because 24 grams of diamond when burned will yield 88 grams of carbon dioxide. O(2)

(C = 12 O = 16)

4.00 27. Addition of barium chloride solution to an unknown solution provides an unmistakable indication of a sulphate because no other class of compounds will form a white precipitate with barium chloride. O(5)

4.00 28. Adding excess sulphur dioxide to a solution of potassium permanganate causes the solution to turn blue because sulphurous acid is a reducing agent. O(4)

4.00 29. The addition of sulphuric acid to sugar results in charring the sugar because the sulphuric acid is a dehydrating agent. O(1)

4.00 30. A piece of filter paper moistened with hot turpentine will burst into flame in chlorine gas because chlorine reacts violently with hydrocarbons to form carbon tetrachloride. O(3)

End of Part A. Go on to Part B

Part B

31. When a candle is burned completely in air, the carbon dioxide and water produced have a weight, compared with that of the candle, which is
    - ○ 1. greater.
    - 2. the same.
    - 3. less.
    - 1.00 4. sometimes greater, sometimes less, depending upon the temperature.
    - 5. Carbon monoxide rather than carbon dioxide is formed by the combustion of the candle.

32. We no longer accept the Phlogiston Theory of Combustion because prediction does not agree with observation in one of the following instances:
    - 4.00 1. Combustion of a substance results in a new substance being formed.
    - 2. New properties may appear when a substance is burned.
    - ○ 3. When a metal is burned, a loss in weight takes place.
    - 4. Metals can be recovered from their ores by heating with charcoal.
    - 5. A substance heated in a limited amount of air is only partially burned.

33. Oxygen may be made by

1.00
1. heating iron oxide in steam.
2. the action of zinc on hydrochloric acid.
3. the action of manganese dioxide on sulphuric acid.
4. the action of hydrochloric acid on calcium carbonate.
○ 5. adding water to sodium peroxide.

34. A glowing splint placed in oxygen will burst into flame. Assume that if placed in other gases lacking in oxygen, the glowing splint is extinguished.

A jar of oxygen and a jar of nitrogen were placed with their mouths separated by a glass plate. The plate was removed for a few seconds, and then a glowing splint was placed in each jar. The splint burst into flame in both containers. This provides direct evidence that

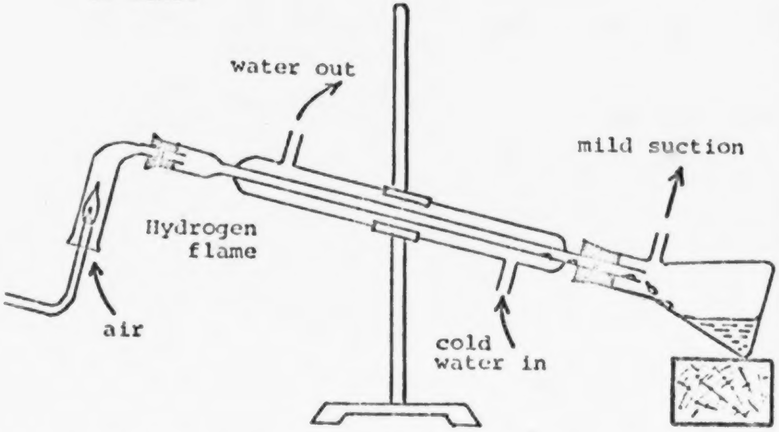
4.00
1. the two gases intermingled.
2. the oxygen diffused into the nitrogen.
3. the nitrogen diffused into the oxygen.
○ 4. oxygen is present in both jars.
5. None of the above is correct.

35. When water is decomposed into its elements, the volume of hydrogen produced is twice the volume of oxygen. The weight of hydrogen produced, compared to the weight of oxygen, is

3.00
1. one-sixteenth as much.
○ 2. one-eighth as much.
3. half as much.
4. twice as much.
5. eight times as much.

36. An element which will not displace hydrogen from acids is

1.00
1. magnesium.
2. sodium.
3. iron.
○ 4. copper.
5. zinc.



37. Imagine the above apparatus to be assembled and operating in the laboratory. This apparatus would be most useful to demonstrate

3.00
1. the analysis of water.
2. the reduction of hydrogen.
○ 3. the synthesis of water.
4. the distillation of water.
5. hydration.

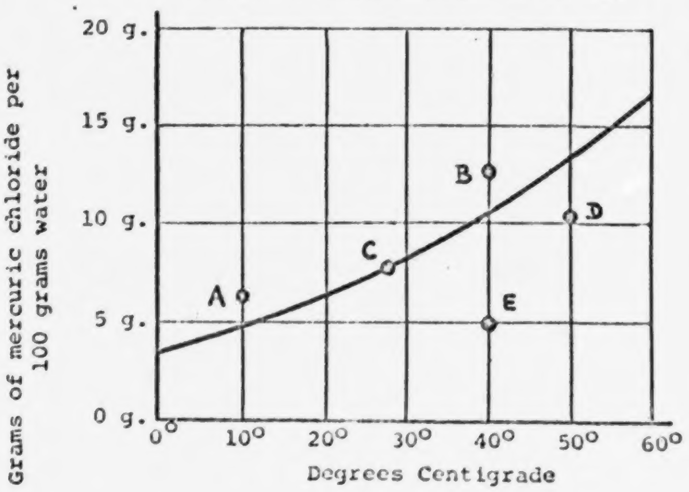
38. A mixture of salt and chalk may be separated into its components by

2.00
1. adding water and shaking.
○ 2. adding water, shaking, filtering and evaporating.
3. adding water and distilling.
4. adding water, boiling and filtering.
5. subliming the salt out of the mixture.

Items 39 to 42:

The following chart is a solubility curve for mercuric chloride dissolved in water. Points A, B, C, D, and E represent five solutions of different compositions at different temperatures.

Use this graph to answer items 39 to 42.



39. Which solution is most dilute ?

2.00
1. A.
2. B.
3. C.
4. D.
○ 5. E.

40. Consider the following statement:

"At 70° C. the solubility would be 10 grams of solute per 50 grams of solvent".

2.00
1. The statement is true.
○ 2. The statement is probably true; additional data would be needed for a final decision.
3. It is impossible to judge the statement because the data are insufficient.
4. The statement is probably false; additional data would be needed for a final decision.
5. The statement is false.

41. About how many grams of water would be required to dissolve 20 grams of mercuric chloride at 40° C ?

3.00
1. 5 grams.
2. 10 grams.
3. 20 grams.
4. 100 grams.
○ 5. 200 grams.

42. A small crystal of mercuric chloride is added to each of the five solutions. Crystallization takes place in

3.00
○ 1. in both A and B.
2. in both A and E.
3. only in C.
4. in both B and D.
5. in both D and E.

43. Consider the following statement and the reason given to support it:

"Dust-free air is considered a solution because dust-free air consists of only one phase, even though its composition may vary somewhat".

○1. Both the statement and reason are true, and the reason supports the statement.
2. Both statement and reason are true, but the reason does not support the statement.
3. The statement is true but the reason is false.
4.00 4. The statement is false but the reason is true.
5. Both the statement and the reason are false.

44. A solid is placed in a liquid and slowly disappears. Which of the following would yield the most reliable evidence in determining whether the solid had merely gone into solution or had actually undergone a chemical change ?

1.00 1. A change in mass.
2. A change in colour.
3. A change in boiling point of the resulting liquid.
○4. Whether or not the substances would separate on evaporation.
5. A change in volume.

45. The chemist Proust once stated: "The cinnabar (mercuric sulphide) of Japan has the same properties and composition as that of Spain. Silver chloride is identical whether obtained in Spain or Peru ". .

This, in effect, is a statement of

2.00 1. The Law of Conservation of Mass.
2. The Law of Multiple Proportions.
○3. The Law of Definite Proportions.
4. The Law of Reacting Volumes.
5. Dulong and Petit's Law.

46. The experimental observation that the volume of a gas may be greatly decreased by applying pressure is best explained by the assumption that

1. the average kinetic energy of gas molecules is directly proportional to the absolute temperature, .
2. gas molecules collide without loss of kinetic energy.
○3. gas molecules are small compared to the distance between them.
4. gas molecules exert almost no attraction on one another.
2.00 5. None of the above explains the observation.

47. One volume of hydrogen reacts with one volume of chlorine to produce two volumes of hydrogen chloride, all substances being gases measured at the same temperature and pressure. Therefore

1. one molecule of each of the elements combines to form one molecule of hydrogen chloride.
2.00 2. each molecule of hydrogen chloride contains two atoms of hydrogen.
3. the volume of the gas produced in any reaction equals the sum of the volumes of the gaseous reactants.
4. the relation is an example of the Law of Combining Weights.
○5. one molecule of each element combines to form two molecules of hydrogen chloride.

48. It is found that two litres of a gas at STP weigh 4.50 grams. The gram-molecular weight of this gas is

1. 4.50 g.
3.00 2. 2.25 g.
3. 100.8 g.
○4. 50.4 g.,
5. none of these.

49. A substance is composed of elements X and Y. Its formula is XY_2 , indicating that

○1. one atom of X combines with two atoms of Y
2. the weight of the element Y in that compound is double that of X.
3. the atomic weight of element Y in the compound is double that of X.
2.00 4. the valence of Y is double that of X.
5. two of the above are correct.

50. Suppose one of the compounds of elements Q and R consists of exactly 25 % Q and 75 % R by weight. What additional information is required in or to calculate the ratio of atoms of Q and R in th compound ?

4.00 1. The molecular weight of the compound.
2. The combining volumes of Q and R involved the formation of the compound.
○3. The relative atomic weights of Q and R.
4. The actual weight of atoms of Q and R.
5. No additional information is needed.

51. 100 grams of calcium carbonate react with hydrochloric acid to form 111 grams of calcium chloride, 18 grams of water and 44 grams of carbon dioxide. The weight of calcium chloride that can be formed from 80 grams of calcium carbonate is about

○1. 89 g.
2. 100 g.
3. 111 g.
3.00 4. 173 g.
5. none of these.

52. When an unknown gas is bubbled through limewater, which is clear and colourless, a white precipitate is formed. The gas can be assumed to be carbon dioxide provided

1. all the following are true.
○2. no other gas forms a white precipitate with limewater.
3. no other substance gives a white precipitate with limewater.
4. that the gas does not react chemically with the limewater.
4.00 5. there is no marked change in the temperature of the limewater.

53. The equation for the manufacture of water gas is

$$H_2O + C \longrightarrow CO + H_2$$

In this reaction carbon is

1. an oxidizing agent

2. a catalyst.

2.00 3. reduced.

○ 4. a reducing agent.

5. a base.

54. Of the following elements, the one that burns with the greatest difficulty is

1.00 1. sulphur.

2. phosphorus.

○ 3. nitrogen.

4. hydrogen.

5. magnesium.

55. Which of the following is the anhydride of H_3PO_4 ?

1. P_2O .

1.00 2. PO_2 .

3. P_2O_3 .

4. PO .

○ 5. P_2O_5 .

56. A mixture of iron and sulphur was heated until no further change occurred. When the product was tested with a magnet, the product was found to be magnetic. A sample of pure iron sulphide was tested with a magnet and found to be non-magnetic.

Which of the following assumptions would explain the magnetic properties of the product ?

1. An excess of sulphur was present in the product.

○ 2. There was not enough sulphur to react with all the iron.

3. There was not enough iron to react with all the sulphur.

4.00 4. Too much heat was applied.

5. The relative amounts of iron and sulphur are not important.

57. Hydrogen sulphide is bubbled into an aqueous solution of arsenic III chloride. The resulting precipitate is

1. reddish-brown.

1.00 2. white.

3. orange.

○ 4. yellow.

5. black.

58. Which of the following is an endothermic reaction ?

○ 1. Heating potassium chlorate to obtain oxygen.

2. Heating magnesium until it catches fire.

3. Adding concentrated sulphuric acid to water.

4. Removing crystals from honey by warming the honey.

1.00 5. Two of the above are endothermic.

59. Metallic sodium is commonly made by

1. heating sodium chloride.

1.00 2. heating sodium bicarbonate.

3. the action of metallic aluminum on sodium chloride.

4. the electrolysis of sodium chloride solution.

○ 5. the electrolysis of molten sodium chloride.

60. Chlorine may be made by

1. the reaction of sulphuric acid on chlorides.

1.00 2. the reaction of manganese dioxide on hydrochloric acid.

3. the electrolysis of brine.

○ 4. two of the above methods.

5. all of the above methods.

End of Test

Go back and check your answers if time permits

APPENDIX G

OTAC Item Analysis

and a

note on variability, relative dispersion, skewness and kurtosis of the test and its subtests.

TABLE G-1

OTAC ITEM RESPONSE PATTERNS

Item No.	Proportion of Subjects Selecting Response					
	Omit	(1)	(2)	(3)	(4)	(5)
1	.007	.102	.203	.027	.138	.524
2	.026	.097	.037	.720	.018	.102
3	.023	.025	.065	.091	.124	.671
4	.059	.148	.554	.159	.042	.037
5	.036	.406	.109	.091	.313	.054
6	.017	.024	.103	.032	.033	.790
7	.072	.094	.119	.469	.076	.171
8	.052	.432	.041	.130	.086	.260
9	.045	.048	.466	.042	.167	.233
10	.106	.174	.050	.094	.339	.238
11	.065	.123	.254	.186	.245	.128
12	.050	.206	.201	.144	.313	.085
13	.035	.097	.062	.579	.076	.149
14	.026	.038	.094	.068	.144	.630
15	.026	.086	.437	.162	.107	.182
16	.024	.047	.032	.706	.153	.037
17	.040	.074	.088	.189	.449	.160
18	.025	.427	.298	.167	.050	.032
19	.052	.375	.111	.133	.086	.244
20	.142	.085	.145	.285	.089	.254
21	.089	.592	.067	.097	.029	.126
22	.039	.008	.050	.101	.744	.058
23	.063	.046	.069	.072	.034	.717
24	.062	.092	.200	.060	.063	.522
25	.040	.245	.131	.316	.131	.138
26	.136	.181	.218	.173	.068	.224
27	.174	.204	.069	.169	.141	.243
28	.136	.054	.084	.200	.287	.239
29	.095	.441	.112	.198	.093	.061
30	.104	.271	.121	.317	.095	.091

TABLE G-1 — Continued

Item No.	Proportion of Subjects Selecting Response					
	Omit	(1)	(2)	(3)	(4)	(5)
31	.059	.114	.434	.266	.062	.065
32	.212	.105	.104	.225	.132	.221
33	.042	.300	.071	.277	.081	.229
34	.019	.177	.111	.050	.474	.169
35	.019	.142	.569	.068	.144	.058
36	.058	.100	.137	.336	.315	.054
37	.055	.074	.173	.239	.318	.141
38	.035	.073	.555	.083	.096	.159
39	.050	.139	.206	.047	.042	.516
40	.090	.070	.229	.275	.134	.201
41	.131	.038	.097	.105	.199	.430
42	.170	.330	.085	.183	.149	.083
43	.065	.233	.152	.142	.206	.202
44	.035	.056	.093	.265	.451	.100
45	.052	.149	.071	.556	.053	.119
46	.078	.047	.032	.425	.032	.386
47	.071	.168	.091	.139	.196	.335
48	.084	.113	.219	.059	.324	.201
49	.044	.535	.028	.031	.059	.303
50	.099	.101	.058	.491	.119	.133
51	.136	.640	.031	.024	.023	.147
52	.068	.314	.458	.068	.047	.045
53	.074	.199	.069	.038	.591	.029
54	.077	.095	.071	.565	.134	.057
55	.254	.043	.206	.163	.115	.219
56	.175	.062	.484	.103	.065	.112
57	.197	.138	.066	.164	.294	.142
58	.163	.117	.076	.168	.077	.398
59	.262	.054	.072	.178	.195	.238
60	.121	.130	.203	.102	.281	.162

TABLE G-2

OTAC ITEM STATISTICS-TAXONOMY CATEGORY 1.00

Item No.	Proportion Answering Correctly	SD	Point-Biserial Correlation	
			Subtest	Total
1	.524	.499	.417	.186
2	.720	.449	.356	.305
3	.671	.470	.326	.277
5	.303	.460	.339	.262
6	.790	.407	.375	.292
7	.469	.499	.383	.338
8	.432	.495	.384	.329
9	.466	.499	.428	.369
12	.206	.404	.322	.278
14	.630	.483	.319	.290
16	.706	.455	.355	.329
17	.449	.497	.344	.296
24	.522	.499	.382	.360
31	.114	.317	.201	.173
33	.229	.420	.331	.275
36	.315	.465	.332	.265
44	.451	.498	.252	.226
54	.565	.496	.391	.331
55	.219	.414	.334	.305
57	.294	.456	.263	.206
58	.117	.321	.214	.155
59	.238	.426	.335	.280
60	.281	.450	.266	.192

TABLE G-2 — Continued

OTAC ITEM STATISTICS-TAXONOMY CATEGORY 2.00

Item No.	Proportion Answering Correctly	SD	Point-Biserial Correlation	
			Subtest	Total
4	.554	.497	.398	.320
13	.579	.494	.382	.283
19	.375	.484	.490	.411
38	.555	.497	.391	.288
39	.516	.500	.475	.368
40	.229	.420	.428	.328
45	.556	.497	.483	.407
46	.425	.494	.475	.379
47	.335	.472	.381	.274
49	.535	.499	.438	.336
53	.591	.492	.438	.382

TABLE G-2 — Continued

OTAC ITEM STATISTICS-TAXONOMY CATEGORY 3.00

Item No.	Proportion Answering Correctly	SD	Point-Biserial Correlation	
			Subtest	Total
10	.339	.473	.315	.287
11	.245	.430	.256	.168
15	.437	.496	.354	.264
18	.298	.457	.350	.277
20	.285	.451	.326	.244
21	.592	.491	.351	.234
22	.744	.436	.402	.275
23	.717	.450	.452	.368
35	.569	.495	.493	.430
37	.239	.427	.391	.367
41	.430	.495	.493	.425
42	.330	.470	.479	.422
48	.324	.468	.458	.381
51	.640	.480	.430	.361

TABLE G-2 — Continued

OTAC ITEM STATISTICS-TAXONOMY CATEGORY 4.00

Item No.	Proportion Answering Correctly	SD	Point-Biserial Correlation	
			Subtest	Total
25	.131	.337	.259	.210
26	.218	.413	.334	.268
27	.243	.429	.302	.136
28	.287	.452	.369	.199
29	.441	.496	.339	.153
30	.317	.465	.349	.175
32	.225	.418	.285	.154
34	.474	.499	.332	.190
43	.233	.423	.332	.254
50	.491	.500	.368	.221
52	.458	.498	.407	.316
56	.484	.500	.410	.363

TABLE G-3

OTAC Summary Statistics

Category	Number of Items	Mean	S.E.	Standard Deviation	S.E.	Skewness	S.E.	Kurtosis	S.E.	Reliability (KR-20)
Total	60	25.151*	.168	8.133*	.122	.481*	.095	.095	.101	.819
1.00	23	9.713*	.072	3.499*	.051	.329*	.051	-.040	.101	.640
2.00	11	5.248*	.048	2.324*	.030	.223*	.051	-.490*	.101	.570
3.00	14	6.190*	.054	2.596*	.034	.178*	.051	-.431*	.101	.590
4.00	12	4.000*	.039	1.869*	.027	.346*	.051	-.103	.101	.315

*Significant at the .001 level.

Note:

Additional descriptive statistics are given in Table 4, pg.131.

Variability and Relative Dispersion

In the present study the term variability is used to identify the dispersion measured by the standard deviation. The term relative dispersion is used to denote the departure from normality measured by the kurtosis of the distribution.

Skewness and Kurtosis

In this study the Fisher statistics g_1 and g_2 (Johnson, 1949, pp.153-158) are used to measure skewness and kurtosis. These statistics are easily interpreted in that a zero value for either statistic indicates no departure from normality. A significantly positive or negative value of g_1 indicates skewness of the same sign as the statistic; a positive value of g_2 indicates a leptokurtic (peaked) distribution and a negative value a platykurtic (flattened) distribution. The g_1 and g_2 statistics also have the advantage of being dimensionless, thus permitting convenient comparison of the dispersions of distributions whose score ranges are not equal. Significant departures from normality are tested by the t test.

Examinations of Table G-3 shows that the total test and all its subtests are positively skewed, an indication that the test as a whole and all its subtests are too difficult for the group tested. The subtests Category 2.00 and Category 3.00 are platykurtic, indicating a tendency towards an excessive number of large deviations from the mean in comparison with a normal distribution of scores.

If the subtests are arranged in order of increasing kurtosis the order of Categories is 2.00, 3.00, 4.00, and 1.00. In other words the scores of Category 2.00 are the most widely scattered over the score range and the scores of Category 1.00 are the least dispersed. The absence of significant leptokurtosis in the OTAC score distributions indicates the absence of heavy concentrations of scores in any specific score range.

APPENDIX H

DESCRIPTIVE STATISTICS

The statistics in this Appendix are presented to describe conditions as they existed in the Spring of 1964, and to facilitate comparison of conditions in Ontario schools with those of other jurisdictions. Means and standard deviations have been transcribed directly from the computer output. The practice of retaining four decimal places has been followed to accommodate other workers who may wish to use these statistics as intermediate steps in their computations.

Selected variables have had their means and standard deviations computed on the bases of school, teacher, class and student.

TABLE H-1

MEANS AND STANDARD DEVIATIONS OF CONTINUOUS STUDENT VARIABLES

Variable	Range of Values	N	Mean	SD
OTAC Total Score	3-56	2,339	25.1509	8.1308
OTAC Category 1.00	0-23	2,339	9.7127	3.4985
OTAC Category 2.00	0-11	2,339	5.2484	2.3238
OTAC Category 3.00	0-14	2,339	6.1898	2.5951
OTAC Category 4.00	0-10	2,339	4.0000	1.8689
SATO Total Verbal	6-52	2,248	26.1824	8.3324
SATO Mathematics	2-30	2,248	17.0525	5.6930
Chemistry Mark, percent	12-100	2,313	60.4228	15.1354
Average Mark, percent	22-96	2,280	62.4399	9.7902
Age nearest birthday, yrs.	15-27	2,210	17.7719	0.9778
Number of older brothers	0-9	2,339	0.4707	0.8542
Number of younger brothers	0-7	2,339	0.7336	1.0111
Number of older sisters	0-7	2,339	0.4408	0.8246
Number of younger sisters	0-9	2,339	0.7174	0.9852
Ordinal position in family	1-14	2,028	2.0513	1.3901
Reverse ordinal position in family	1-11	2,028	2.6736	1.6312
Number of children in family	1-18	2,230	3.4780	1.9365

TABLE H-2

MEDIANS, MEANS AND STANDARD DEVIATIONS OF CONTINUOUS TEACHER VARIABLES

Variable	Number	Frequency	Statistic
Grade 12 Chemistry Teaching Experience in years	1	11	Median = 3.5
	2	4	
	3	9	Mean = 8.4792
	4	3	
	6	3	SD = 9.7979
	7	1	
	8	2	N = 48
	9	1	
	10	1	
	11	1	
	13	1	
	16	2	
	18	1	
	19	1	
	20	2	
	22	1	
	27	2	
	38	1	
	42	1	
Grade 13 Chemistry Teaching Experience in years	0	20	Median = 1.5
	1	4	
	2	4	Mean = 4.6458
	3	9	
	4	2	SD = 8.3927
	5	1	
	12	1	N = 48
	13	1	
	15	1	
	18	2	
	27	1	
	30	1	
	38	1	

TABLE H-2 — Continued

Variable	Number	Frequency	Statistic
Number of class preparations per week	10	1	Median = 21.5
	13	1	
	14	2	Mean = 22.2917
	15	6	
	16	1	SD = 6.6394
	17	2	
	18	1	N = 48
	20	5	
	21	5	
	22	2	
	23	3	
	24	5	
	25	2	
	26	1	
	27	1	
	28	3	
	29	2	
	30	1	
	32	1	
	35	1	
	36	1	
	45	1	
Number of Grade 12 Chemistry classes per week	3	1	Median = 6
	5	12	
	6	12	Mean = 9.2917
	10	10	
	12	4	SD = 5.0825
	13	1	
	15	2	N = 48
	18	4	
	24	2	

TABLE H-2 — Continued

Variable	Number	Frequency	Statistic
Number of Grade 13 Chemistry classes per week	0	16	Median = 5
	4	1	
	5	8	Mean = 6.1667
	6	7	
	8	2	SD = 6.0530
	9	1	
	10	3	N = 48
	11	1	
	12	3	
	14	1	
	15	2	
	16	1	
	24	2	
Number of teaching periods per week	13	1	Median = 36
	15	1	
	16	1	Mean = 34.1458
	25	3	
	26	1	SD = 7.0297
	29	1	
	30	6	N = 48
	31	1	
	34	3	
	35	4	
	36	7	
	37	3	
	38	3	
	39	2	
	40	3	
	41	4	
	42	2	
	44	1	
	45	1	

TABLE H-2 — Continued

Variable	Number	Frequency	Statistic
Teaching time per week, hours	9	2	Median = 21.5
	10	1	
	15	1	Mean = 21.1459
	16.5	3	
	17	1	SD = 4.0952
	20	5	
	20.5	2	N = 48
	21	7	
	21.5	4	
	22	4	
	22.5	1	
	23.5	4	
	24	3	
	24.5	4	
	25	1	
	25.5	2	
	26	2	
	29	1	
Total number of pupils per week	41-60	1	Median =180
	61-80	0	
	81-100	3	Mean =180.4348
	101-120	2	
	121-140	6	SD = 37.9477
	141-160	2	
	161-180	11	N = 46
	181-200	4	
	201-220	6	
	221-240	5	
	241-260	5	
	261-280	1	
	Not reported	2	

TABLE H-3

MEANS AND STANDARD DEVIATIONS OF CONTINUOUS SCHOOL VARIABLES COMPUTED ON VARIOUS BASES

Variable	Range	Basis	N	Mean	SD
Size of Chemistry Class	15-41	by school	30	27.7667	5.6785
		by teacher	48	28.6250	5.6555
		by class	80	30.4375	6.0619
		by student	2227	31.5905	5.6017
Length of Class Period for Chemistry, min.	33-45	by school	30	37.4667	3.0412
		by teacher	48	37.6667	3.1380
		by class	80	37.5125	2.8896
		by student	2339	37.6139	2.9837
Number of class periods per week allotted to Grade 12 Chemistry	3-7	by school	30	5.4000	0.6633
		by teacher	48	5.4375	0.6092
		by class	80	5.4875	0.5700
		by student	2339	5.5019	0.5536

Note:

Interpretation of the first variable is as follows:
The average school had 27.8 students per class.
The average teacher had 28.6 students per class.
The average class had 30.4 students.
The average student was in a class of 31.6 students.

Similar interpretations are made for the other variables.

TABLE H-4

FREQUENCY DISTRIBUTIONS, MEANS AND STANDARD DEVIATIONS OF INVENTORY OF CHOICES 4-POINT SCALE SCORES

N = 2339

Scale	Score[a]	Frequency	Mean[b]	SD[b]
Prudent-Theoretic	0	291	1.5952	0.9702
	1	649		
	2	661		
	3	415		
	E	323		
Prudent-Immediate	0	211	1.6757	0.9572
	1	589		
	2	594		
	3	422		
	E	523		
Prudent-Aesthetic	0	557	0.9847	0.8082
	1	981		
	2	327		
	3	100		
	E	374		
Theoretic-Immediate	0	341	1.3745	0.9028
	1	721		
	2	630		
	3	212		
	E	435		
Theoretic-Aesthetic	0	418	1.3838	0.9676
	1	571		
	2	657		
	3	243		
	E	450		
Aesthetic-Immediate	0	133	1.6425	0.9100
	1	806		
	2	480		
	3	416		
	E	504		

[a] The category E includes "error" scores and students who did not write the Inventory of Choices.

[b] Category E students not included in computations.

TABLE II-5

FREQUENCY DISTRIBUTIONS, MEANS AND STANDARD DEVIATIONS OF INVENTORY OF CHOICES 9-POINT SCALE SCORES

N = 2339

Scale	Score[a]	Frequency	Mean[b]	SD[b]
Prudent-Theoretic	0	205	4.4902	2.0561
	1	151		
	2	245		
	3	392		
	4	221		
	5	467		
	6	324		
	7	163		
	8	78		
	9	93		
Prudent-Immediate	0	236	4.1793	1.9748
	1	221		
	2	283		
	3	314		
	4	328		
	5	363		
	6	334		
	7	167		
	8	77		
	9	16		
Prudent-Aesthetic	0	192	4.3940	1.9183
	1	159		
	2	240		
	3	336		
	4	374		
	5	362		
	6	378		
	7	202		
	8	65		
	9	31		
Theoretic-Immediate	0	179	4.6958	1.9294
	1	97		
	2	206		
	3	338		
	4	339		
	5	436		
	6	351		
	7	209		
	8	141		
	9	43		

TABLE H-5 — Continued

Scale	Score[a]	Frequency	Mean[b]	SD[b]
Theoretic-Aesthetic	0	202	4.3921	2.1335
	1	263		
	2	212		
	3	291		
	4	316		
	5	357		
	6	307		
	7	223		
	8	138		
	9	30		
Aesthetic-Immediate	0	143	4.5544	2.0659
	1	181		
	2	193		
	3	339		
	4	395		
	5	371		
	6	301		
	7	213		
	8	140		
	9	63		

[a] Scores of "0" include those students with insufficient responses to score and those not writing the Inventory of Choices.

[b] Scores of "0" not included in computations.

TABLE H-6

FREQUENCY DISTRIBUTIONS, MEANS AND STANDARD DEVIATIONS OF INVENTORY OF CHOICES 12-POINT SCALE SCORES

N = 2339

Scale	Score[a]	Frequency	Mean[b]	SD[b]
Prudent-Theoretic	0	155	6.4162	2.9581
	1	84		
	2	188		
	3	154		
	4	179		
	5	296		
	6	139		
	7	353		
	8	198		
	9	237		
	10	139		
	11	121		
	12	96		
Prudent-Immediate	0	194	5.5268	2.6249
	1	116		
	2	185		
	3	244		
	4	284		
	5	277		
	6	265		
	7	214		
	8	225		
	9	194		
	10	82		
	11	46		
	12	13		
Prudent-Aesthetic	0	267	4.7156	2.5400
	1	257		
	2	211		
	3	397		
	4	266		
	5	238		
	6	245		
	7	232		
	8	165		
	9	107		
	10	25		
	11	27		
	12	2		

TABLE H-6 — Continued

Scale	Score[a]	Frequency	Mean[b]	SD[b]
Theoretic-Immediate	0	131	6.4040	2.8381
	1	92		
	2	126		
	3	181		
	4	223		
	5	226		
	6	272		
	7	285		
	8	200		
	9	240		
	10	201		
	11	110		
	12	52		
Theoretic-Aesthetic	0	146	5.4295	2.6323
	1	143		
	2	206		
	3	226		
	4	267		
	5	291		
	6	284		
	7	312		
	8	164		
	9	140		
	10	88		
	11	53		
	12	19		
Aesthetic-Immediate	0	127	6.3395	2.5756
	1	96		
	2	57		
	3	172		
	4	225		
	5	242		
	6	350		
	7	348		
	8	289		
	9	164		
	10	139		
	11	87		
	12	43		

[a] Scores of "0" include those students with insufficient responses to score and those not writing the Inventory of Choices.

[b] Scores of "0" not included in computations.

TABLE H-7

FREQUENCY DISTRIBUTIONS OF CATEGORICAL STUDENT VARIABLES

Variable	Code	Interpretation	Frequency
Sex of student	0	not indicated[a]	116
	1	male	1173
	2	female	1050
Language spoken in the home	0	not indicated	129
	1	English	2003
	2	French	13
	3	other	194
Length of residence in Ontario of family	0	not indicated	142
	1	less than 1 year	7
	2	less than 2 years	7
	3	less than 5 years	22
	4	less than 10 years	134
	5	10 - 24 years	499
	6	25 - 49 years	407
	7	50 - 74 years	349
	8	75 - 99 years	240
	9	over 100 years	532
Occupation of father[b]	0	not indicated	108
	1	Manual labor	378
	2	Skilled manual labor	315
	3	Lower white collar	326
	4	Upper white collar	389
	5	Small business (s.e.)[c]	126
	6	Merchants (s.e.)	72
	7	Farmers and ranchers (s.e.)	258
	8	Professional (salaried)	84
	9	Professional (s.e.)	60
	10	Executive, Vice-president, VIP's	61
	11	Farm laborer	2
	12	Unemployed, retired, deceased	160
Occupation of mother[b]	0	not indicated	108
	1	Manual labor	93
	2	Skilled manual labor	1
	3	Lower white collar	339
	4	Upper white collar	186
	5	Small business (s.e.)	9
	6	Merchants (s.e.)	13
	7	Farmers and ranchers (s.e.)	1

TABLE H-7 --- Continued

Variable	Code	Interpretation	Frequency
Occupation of mother (continued)	8	Professional (salaried)	2
	9	Professional (s.e.)	1
	10	Executive, Vice-president, VIP's	1
	11	Housewife	1447
	12	Unemployed, retired, deceased	138
Occupational Aspiration of Student[b]	0	not indicated	108
	1	Manual labor	9
	2	Skilled manual labor	29
	3	Lower white collar	193
	4	Upper white collar	1047
	5	Small business (s.e.)	16
	6	Merchants (s.e.)	5
	7	Farmers and ranchers (s.e.)	30
	8	Professional (salaried)	235
	9	Professional (s.e.)	215
	10	Executive, Vice-president, VIP's	30
	11	Housewife	143
	12	not certain	279
Subject liked best	0	not indicated	123
	1	English	350
	2	Foreign Language	281
	3	History and Geography	371
	4	Mathematics	473
	5	Science	401
	6	Commercial Subjects	59
	7	Industrial Arts or Home Economics	178
	8	Art	39
	9	Music	64
Subject liked least	0	not indicated	148
	1	English	230
	2	Foreign Language	522
	3	History and Geography	441
	4	Mathematics	386
	5	Science	172
	6	Commercial subjects	155
	7	Industrial Arts or Home Economics	109
	8	Art	74
	9	Music	102

TABLE H-7 — Continued

Variable	Code	Interpretation	Frequency
Repeating Grade 12 Chemistry	0	not indicated	118
	1	no	2004
	2	yes	217
Attitude toward school	0	not indicated	123
	1	like very much	597
	2	like somewhat	1087
	3	like slightly	270
	4	dislike slightly	136
	5	dislike somewhat	83
	6	dislike very mcuh	43
Educational Plans No. 1 (Immediate)	0	not indicated	119
	1	complete Grade 12 only	591
	2	complete Grade 13	1496
	3	leave before completing Grade 12	6
	4	undecided	127
Educational Plans No. 2 (Future)	0	not indicated	122
	1	enter university	963
	2	enter teacher's college[d]	250
	3	enter nursing	188
	4	enter technical or trade training	171
	5	enter business college	93
	6	obtain a job	148
	7	work at home	9
	8	other plans	111
	9	undecided	304

[a] In this table the category "not indicated" includes 108 students in the sample who did not answer the Personal Information Questionnaire.

[b] Categories 11 and 12 were treated as missing data in the Automatic Interaction Detector analysis.

[c] Self-employed.

[d] In Ontario, Teachers' Colleges train elementary school teachers; Colleges of Education train secondary school teachers. For teachers of academic subjects, a bachelor's degree is a prerequisite for entrance to a College of Education.

TABLE H-8

FREQUENCY DISTRIBUTIONS OF CATEGORICAL VARIABLES CLASSIFIED BY TEACHER AND BY STUDENT

Variable	Code	Interpretation	Teacher Frequency N = 48	Student Frequency N = 2339
Chemistry Teacher's Level of Responsibility	0	not indicated	--	112[a]
	1	instructor	31	1519
	2	department head	13	601
	3	vice-principal or principal	4	107
Sex of Teacher	0	not indicated	--	112[a]
	1	male	39	1808
	2	female	9	419
Teacher Qualification[b]	0	not applicable[c]	6	293
	1	OSSTF Category 1	9	471
	2	OSSTF Category 2	11	514
	3	OSSTF Category 3	4	222
	4	OSSTF Category 4	18	839
Textbook used	1	Croal, Couke & Louden (1958)	34	1746
	2	Cragg, Graham & Young (1959)	14	593
Laboratory Manual used[d]	1	Couke and Louden	27	1360
	2	Motherwell and Young	13	493
	3	Herron and Maddeford	2	138
	4	Clemence	4	204
	5	No printed manual used	2	144

TABLE H-8 — Continued

Variable	Code	Interpretation	Teacher Frequency N = 48	Student Frequency N = 2339
Use of Audiovisual aids, programmed instruction, and other instructional aids	0	not indicated	--	112[a]
	1	none	27	1404
	2	conventional films	6	185
	3	films and models	2	43
	4	films and film strips	1	30
	5	films, models, tours, speakers	1	54
	6	programmed instruction and films	10	429
	7	CHEM Study films	1	82

[a] Some students did not answer the personal information questionnaire and others did not name their chemistry teacher.

[b] The Ontario Secondary School Teachers' Federation (OSSTF) has classified teachers in four qualification groups. Category 1 represents the lowest certifiable qualification and category 4 the highest.

[c] Private and independent schools giving work at the high school level are not required to employ certificated teachers; some of the teachers of these schools are therefore not assigned an OSSTF category.

[d] See bibliographical list of laboratory manuals on next page.

TABLE H-8 — Continued

[d] 1. Couke J.H., and Louden A.H. *Experiments in Laboratory Chemistry*. Revised edition. Toronto: Copp Clark Publishing Co. Ltd., 1959.

2. Motherwell, G.W., and Young J.V. *The Elements of Chemistry in the Laboratory*. Toronto: Clarke, Irwin and Company Ltd., 1959.

3. Herron, Walter M., and Maddeford, Charles W. *A First Laboratory Manual in Chemistry*. Toronto: Longmans Canada Ltd., 1964.

4. Clemence, A.L. *A Program Laboratory Course of Study in Chemistry*. Toronto: Service and Smiles Distributors Ltd., n.d.

TABLE H-9

SCHOOLS PARTICIPATING IN THE STUDY

	School	Type[a]	Number of Chemistry Teachers	Number of Chemistry Students Participating
011	Banting Memorial High School, Alliston	1	3	97
037	St. Joseph's Private High School, Barry's Bay	2	1	16
080	Burlington Central High School, Burlington	1	2	178
157	Lambton-Kent District High School, Dresden	1	2	79
158	Dryden High School, Dryden	1	3	94
214	Gananoque High School, Gananoque	1	1	86
248	Cathedral Girls' High School, Hamilton	2	1	115
251	Hillfield-Strathallan Colleges, Hamilton	3	1	22
331	Lively High School, Lively	1	1	44
510	Lakeview High School, Port Arthur	1	2	89
521	Saugeen District High School, Port Elgin	1	1	55
532	Preston High School, Preston	1	3	76
535	Rainy River High School, Rainy River	1	1	24
609	Schreiber High School, Schreiber	1	1	13
612	Seaforth High School, Seaforth	1	1	48
627	South Lincoln District High School, Smithville	1	1	29
663	Sydenham High School, Sydenham	1	1	88
715	Humberside Collegiate Institute, Toronto	1	4	204
748	St. Joseph's Private School, Islington	2	2	93

TABLE H-9 — Continued

	School	Type[a]	Number of Chemistry Teachers	Number of Chemistry Students Participating
751	St. Michael's College School, Toronto	2	3	187
794	Notre Dame Academy, Waterdown	2	1	25
804	Ontario Ladies' College, Whitby	3	1	23
841	Huron Park Secondary School, Woodstock	1	3	117
922	Westgate Collegiate and Vocational Institute, Fort William	1	1	100
931	Anderson High School, Whitby	1	1	60
946	St. Thomas Collegiate Institute, St. Thomas	1	1	95
959	Stayner Collegiate Institute, Stayner	1	1	27
986	Midland Avenue Collegiate Institute, Scarborough	1	2	152
988	North Dundas District High School, Chesterville	1	1	82

[a] 1. Publicly supported high schools
2. Roman Catholic Private schools
3. Independent Private schools

TABLE H-10

SUMMARY OF SCHOOLS PARTICIPATING IN THE STUDY

Type of School	Number of Schools	Number of Chemistry Teachers	Number of Classes	Number of Students
1. Publicly supported high schools	22	36	63	1837
2. Roman Catholic Private schools	6	10	15	457
3. Independent Private schools	2	2	2	45
Totals	30	48	80	2339

APPENDIX I

Pattern Analysis Tables

This Appendix contains two-way analysis of variance tables for all groups having congruent or parallel patterns extending over the complete range of a classificatory variable. The sums of squares entries usually found in such tables are omitted. While entries in the tables are expressed to four decimal places, values of F and R_p, when used in the text, have been rounded off to two decimal places.

With the exception of Tables I-35, I-36, and I-37, the analyses concern patterns consisting only of scores in Categories 2.00, 3.00 and 4.00 of the Taxonomy.

For comparison, four groups not having congruent or parallel patterns have been included. Tables I-34, I-35 and I-36 illustrate small groups whose interaction F's are significant. (Table I-3 is an example of a large group having a non-significant F_{ck}). Tables I-36 and I-37 illustrate analyses of patterns composed of four Taxonomy category scores.

Critical values of F for high values of df_1 and df_2 were calculated using the formula given by Dixon and Massey (1957, p.402).

TABLE I-1

ANALYSIS OF AID GROUP ED PLANS 1 STUDENTS NOT INTENDING TO COMPLETE GRADE 13

Subtest	Mean Stabilized Score	Grand Mean
2.00	7.0688	6.5468
3.00	7.1089	
4.00	5.4627	

Summary of Findings

Source of Variation	df.	Mean Square	F	p	R_p	p
Subtests	2	380.0825	314.8203	.001		
Students	430	2.0724	2.0724	.001		
Subtests x Students	860	1.2073	1.2073		.828	.05
Residual	(182)	(1.00)				

TABLE I-2

ANALYSIS OF AID GROUP ED PLANS 1 STUDENTS INTENDING TO COMPLETE GRADE 13

Subtest	Mean Stabilized Score	Grand Mean
2.00	7.3043	6.7957
3.00	7.4307	
4.00	5.6520	

Summary of Findings

Source of Variation	df.	Mean Square	F	p	R_p	p
Subtests	2	412.6905	343.7083	.001		
Students	418	2.3730	2.3730	.001		
Subtests x Students	836	1.2007	1.2007		.833	.05
Residual	(182)	(1.00)				

TABLE I-3

ANALYSIS OF AID GROUP ED PLANS 1
ALL STUDENTS COMBINED

Subtest	Mean Stabilized Score	Grand Mean
2.00	7.1849	6.6695
3.00	7.2675	
4.00	5.5560	

Summary of Findings

Source of Variation	df.	Mean Square	F	p	R_p	p
Subtests	2	791.8130	657.7613	.001		
Students	849	2.2645	2.2645	.001		
Subtests x Students	1698	1.2038	1.2038		.831	.05
Residual	(182)	(1.00)				

TABLE I-4

ANALYSIS OF HIGHLY IMMEDIATE UNDERACHIEVERS
THEORETIC-IMMEDIATE SCORE OF 0

Subtest	Mean Stabilized Score	Grand Mean
2.00	5.7740	5.2843
3.00	5.4483	
4.00	4.6305	

Summary of Findings

Source of Variation	df.	Mean Square	F	p	R_p	p
Subtests	2	19.4361	16.9113	.001		
Students	55	1.2150	1.2150			
Subtests x Students	110	1.1493	1.1493		.870	.05
Residual	(182)	(1.00)				

TABLE I-5

ANALYSIS OF MODERATELY IMMEDIATE UNDERACHIEVERS
THEORETIC-IMMEDIATE SCORE OF 1

Subtest	Mean Stabilized Score	Grand Mean
2.00	5.6206	5.3212
3.00	5.7379	
4.00	4.6051	

Summary of Findings

Source of Variation	df.	Mean Square	F	p	R_p	p
Subtests	2	43.4597	38.3852	.001		
Students	111	1.1802	1.1802			
Subtests x Students	222	1.1322	1.1322		.883	.05
Residual	(182)	(1.00)				

TABLE I-6

ANALYSIS OF MODERATELY THEORETIC UNDERACHIEVERS THEORETIC-IMMEDIATE SCORE OF 2

Subtest	Mean Stabilized Score	Grand Mean
2.00	5.5921	5.3990
3.00	5.9665	
4.00	4.6384	

Summary of Findings

Source of Variation	df.	Mean Square	F	p	R_p	p
Subtests	2	27.6644	25.1152	.001		
Students	58	0.8606	0.8606			
Subtests x Students	116	1.1015	1.1015		.908	.05
Residual	(182)	(1.00)				

TABLE I-7

ANALYSIS OF HIGHLY THEORETIC UNDERACHIEVERS THEORETIC-IMMEDIATE SCORE OF 3

Subtest	Mean Stabilized Score	Grand Mean
2.00	5.5348	5.4514
3.00	6.1092	
4.00	4.7102	

Summary of Findings

Source of Variation	df.	Mean Square	F	p	R_p	p
Subtests	2	10.3856	8.2774	.001		
Students	20	0.9061	0.9061			
Subtests x Students	40	1.2547	1.2547		.797	.05
Residual	(182)	(1.00)				

TABLE I-8

ANALYSIS OF HIGHLY IMMEDIATE OVERACHIEVERS THEORETIC-IMMEDIATE SCORE OF 0

Subtest	Mean Stabilized Score	Grand Mean
2.00	8.5220	7.9221
3.00	8.4041	
4.00	6.8401	

Summary of Findings

Source of Variation	df.	Mean Square	F	p	R_p	p
Subtests	2	22.9196	17.5629	.001		
Students	25	0.6958	0.6958			
Subtests x Students	50	1.3050	1.3050		.766	.05
Residual	(182)	(1.00)				

TABLE I-9

ANALYSIS OF MODERATELY IMMEDIATE OVERACHIEVERS THEORETIC-IMMEDIATE SCORE OF 1

Subtest	Mean Stabilized Score	Grand Mean
2.00	8.6820	8.0275
3.00	8.9110	
4.00	6.4896	

Summary of Findings

Source of Variation	df.	Mean Square	F	p	R_p	p
Subtests	2	166.1903	157.2580	.001		
Students	92	0.9649	0.9649			
Subtests x Students	184	1.0568	1.0568		.946	.05
Residual	(182)	(1.00)				

TABLE I-10

ANALYSIS OF MODERATELY THEORETIC OVERACHIEVERS
THEORETIC-IMMEDIATE SCORE OF 2

Subtest	Mean Stabilized Score	Grand Mean
2.00	8.5882	8.0027
3.00	8.8635	
4.00	6.5564	

Summary of Findings

Source of Variation	df.	Mean Square	F	p	R_p	p
Subtests	2	158.7854	130.3657	.001		
Students	99	0.9904	0.9904			
Subtests x Students	198	1.2180	1.2180		.821	.05
Residual	(182)	(1.00)				

TABLE I-11

ANALYSIS OF HIGHLY THEORETIC OVERACHIEVERS
THEORETIC-IMMEDIATE SCORE OF 3

Subtest	Mean Stabilized Score	Grand Mean
2.00	8.7164	8.1086
3.00	9.0180	
4.00	6.5913	

Summary of Findings

Source of Variation	df.	Mean Square	F	p	R_p	p
Subtests	2	101.4594	99.0524	.001		
Students	57	1.0442	1.0442			
Subtests x Students	114	1.0243	1.0243		.976	.05
Residual	(182)	(1.00)				

TABLE I-12

ANALYSIS OF IMMEDIATE UNDERACHIEVERS THEORETIC-IMMEDIATE SCORES OF 0 AND 1

Subtest	Mean Stabilized Score	Grand Mean
2.00	5.6717	5.3089
3.00	5.6414	
4.00	4.6136	

Summary of Findings

Source of Variation	df.	Mean Square	F	p	R_p	p
Subtests	2	60.9546	53.3426	.001		
Students	167	1.1855	1.1855			
Subtests x Students	334	1.1427	1.1427		.875	.05
Residual	(182)	(1.00)				

TABLE I-13

ANALYSIS OF THEORETIC UNDERACHIEVERS THEORETIC-IMMEDIATE SCORES OF 2 AND 3

Subtest	Mean Stabilized Score	Grand Mean
2.00	5.5770	5.4128
3.00	6.0040	
4.00	4.6573	

Summary of Findings

Source of Variation	df.	Mean Square	F	p	R_p	p
Subtests	2	37.8907	33.5791	.001		
Students	79	0.8628	0.8628			
Subtests x Students	158	1.1284	1.1284		.886	.05
Residual	(182)	(1.00)				

TABLE I-14

ANALYSIS OF IMMEDIATE OVERACHIEVERS THEORETIC-IMMEDIATE SCORES OF 0 AND 1

Subtest	Mean Stabilized Score	Grand Mean
2.00	8.6471	8.0045
3.00	8.8002	
4.00	6.5662	

Summary of Findings

Source of Variation	df.	Mean Square	F	p	R_p	p
Subtests	2	185.3311	163.6622	.001		
Students	118	0.9054	0.9054			
Subtests x Students	236	1.1324	1.1324		.883	.05
Residual	(182)	(1.00)				

TABLE I-15

ANALYSIS OF THEORETIC OVERACHIEVERS THEORETIC-IMMEDIATE SCORES OF 2 AND 3

Subtest	Mean Stabilized Score	Grand Mean
2.00	8.6353	8.0416
3.00	8.9202	
4.00	6.5692	

Summary of Findings

Source of Variation	df.	Mean Square	F	p	R_p	p
Subtests	2	260.0999	227.9978	.001		
Students	157	1.0115	1.0115			
Subtests x Students	314	1.1408	1.1408		.877	.05
Residual	(182)	(1.00)				

TABLE I-16

ANALYSIS OF ALL UNDERACHIEVERS HAVING A THEORETIC-IMMEDIATE SCORE

Subtest	Mean Stabilized Score	Grand Mean
2.00	5.6412	5.3424
3.00	5.7584	
4.00	4.6277	

Summary of Findings

Source of Variation	df.	Mean Square	F	p	R_p	p
Subtests	2	95.8653	83.6886	.001		
Students	247	1.0846	1.0846			
Subtests x Students	494	1.1455	1.1455		.873	.05
Residual	(182)	(1.00)				

TABLE I-17

ANALYSIS OF ALL OVERACHIEVERS HAVING A THEORETIC-IMMEDIATE SCORE

Subtest	Mean Stabilized Score	Grand Mean
2.00	8.6403	8.0256
3.00	8.8687	
4.00	6.5679	

Summary of Findings

Source of Variation	df.	Mean Square	F	p	R_p	p
Subtests	2	445.0771	392.3458	.001		
Students	276	0.9635	0.9635			
Subtests x Students	552	1.1344	1.1344		.882	.05
Residual	(182)	(1.00)				

TABLE I-18

ANALYSIS OF ALL UNDERACHIEVERS

Subtest	Mean Stabilized Score	Grand Mean
2.00	5.6433	5.3121
3.00	5.7286	
4.00	4.5644	

Summary of Findings

Source of Variation	df.	Mean Square	F	p	R_p	p
Subtests	2	131.3918	109.4840	.001		
Students	311	1.1555	1.1555			
Subtests x Students	622	1.2001	1.2001		.833	.05
Residual	(182)	(1.00)				

TABLE I-19

ANALYSIS OF ALL OVERACHIEVERS

Subtest	Mean Stabilized Score	Grand Mean
2.00	8.6849	8.0340
3.00	8.8658	
4.00	6.5513	

Summary of Findings

Source of Variation	df.	Mean Square	F	p	R_p	p
Subtests	2	558.3978	487.5134	.001		
Students	336	0.9315	0.9315			
Subtests x Students	672	1.1454	1.1454		.873	.05
Residual	(182)	(1.00)				

TABLE I-20

ANALYSIS OF ALL UNDERACHIEVERS AND OVERACHIEVERS COMBINED

Subtest	Mean Stabilized Score	Grand Mean
2.00	7.2227	6.7255
3.00	7.3576	
4.00	5.5961	

Summary of Findings

Source of Variation	df.	Mean Square	F	p	R_p	p
Subtests	2	623.7824	490.5107	.001		
Students	648	6.5945	6.5945	.001		
Subtests x Students	1296	1.2717	1.2717	.025		
Residual	(182)	(1.00)				

TABLE I-21

ANALYSIS OF ALL UNDERACHIEVERS AND OVERACHIEVERS HAVING A THEORETIC-IMMEDIATE SCORE

Subtest	Mean Stabilized Score	Grand Mean
2.00	7.2236	6.7581
3.00	7.3994	
4.00	5.6514	

Summary of Findings

Source of Variation	df.	Mean Square	F	p	R_p	p
Subtests	2	486.3600	391.7204	.001		
Students	524	6.4123	6.4123	.001		
Subtests x Students	1048	1.2416	1.2416	.05		
Residual	(182)	(1.00)				

TABLE I-22

ANALYSIS OF HIGHLY IMMEDIATE NON-NORMAL ACHIEVERS
THEORETIC-IMMEDIATE SCORE OF 0

Subtest	Mean Stabilized Score	Grand Mean
2.00	6.6454	6.1207
3.00	6.3855	
4.00	5.3311	

Summary of Findings

Source of Variation	df.	Mean Square	F	p	R_p	p
Subtests	2	39.7216	32.6739	.001		
Students	81	5.6155	5.6155	.001		
Subtests x Students	162	1.2157	1.2157		.823	.05
Residual	(182)	(1.00)				

TABLE I-23

ANALYSIS OF MODERATELY IMMEDIATE NON-NORMAL ACHIEVERS
THEORETIC-IMMEDIATE SCORE OF 1

Subtest	Mean Stabilized Score	Grand Mean
2.00	7.0094	6.5490
3.00	7.1774	
4.00	5.4600	

Summary of Findings

Source of Variation	df.	Mean Square	F	p	R_p	p
Subtests	2	183.7542	150.6676	.001		
Students	204	6.5499	6.5499	.001		
Subtests x Students	408	1.2196	1.2196		.820	.05
Residual	(182)	(1.00)				

TABLE I-24

ANALYSIS OF MODERATELY THEORETIC NON-NORMAL ACHIEVERS THEORETIC-IMMEDIATE SCORE OF 2

Subtest	Mean Stabilized Score	Grand Mean
2.00	7.4764	7.0365
3.00	7.7885	
4.00	5.8447	

Summary of Findings

Source of Variation	df.	Mean Square	F	p	R_p	p
Subtests	2	173.2710	138.5171	.001		
Students	158	5.7128	5.7128	.001		
Subtests x Students	316	1.2509	1.2509	.05		
Residual	(182)	(1.00)				

TABLE I-25

ANALYSIS OF HIGHLY THEORETIC NON-NORMAL ACHIEVERS THEORETIC-IMMEDIATE SCORE OF 3

Subtest	Mean Stabilized Score	Grand Mean
2.00	7.8707	7.4022
3.00	8.2448	
4.00	6.0913	

Summary of Findings

Source of Variation	df.	Mean Square	F	p	R_p	p
Subtests	2	104.5940	89.9192	.001		
Students	78	5.1824	5.1824	.001		
Subtests x Students	156	1.1632	1.1632		.860	.05
Residual	(182)	(1.00)				

TABLE I-26

ANALYSIS OF NON-NORMAL ACHIEVERS WITH
EXTREME THEORETIC-IMMEDIATE SCORES

Subtest	Mean Stabilized Score	Grand Mean
2.00	7.2466	6.7495
3.00	7.2978	
4.00	5.7041	

Summary of Findings

Source of Variation	df.	Mean Square	F	p	R_p	p
Subtests	2	132.0682	104.8910	.001		
Students	160	6.6083	6.6083	.001		
Subtests x Students	320	1.2591	1.2591	.05		
Residual	(182)	(1.00)				

TABLE I-27

ANALYSIS OF HIGHLY THEORETIC OVERACHIEVERS
PRUDENT-THEORETIC SCORE OF 0

Subtest	Mean Stabilized Score	Grand Mean
2.00	8.8490	8.0386
3.00	8.7562	
4.00	6.5106	

Summary of Findings

Source of Variation	df.	Mean Square	F	p	R_p	p
Subtests	2	138.5039	137.1869	.001		
Students	78	0.8879	0.8879			
Subtests x Students	156	1.0096	1.0096		.990	.05
Residual	(182)	(1.00)				

TABLE I-28

ANALYSIS OF MODERATELY THEORETIC OVERACHIEVERS
PRUDENT-THEORETIC SCORE OF 1

Subtest	Mean Stabilized Score	Grand Mean
2.00	8.6597	8.0788
3.00	8.9508	
4.00	6.6260	

Summary of Findings

Source of Variation	df.	Mean Square	F	p	R_p	p
Subtests	2	154.0101	142.8268	.001		
Students	95	0.9225	0.9225			
Subtests x Students	190	1.0783	1.0783		.927	.05
Residual	(182)	(1.00)				

TABLE I-29

ANALYSIS OF MODERATELY PRUDENT OVERACHIEVERS
PRUDENT-THEORETIC SCORE OF 2

Subtest	Mean Stabilized Score	Grand Mean
2.00	8.3389	7.9293
3.00	8.8580	
4.00	6.5910	

Summary of Findings

Source of Variation	df.	Mean Square	F	p	R_p	p
Subtests	2	114.2668	95.9741	.001		
Students	80	0.8264	0.8264			
Subtests x Students	160	1.1906	1.1906		.840	.05
Residual	(182)	(1.00)				

TABLE I-30

ANALYSIS OF HIGHLY PRUDENT OVERACHIEVERS
PRUDENT-THEORETIC SCORE OF 3

Subtest	Mean Stabilized Score	Grand Mean
2.00	8.6359	7.9237
3.00	8.7969	
4.00	6.3384	

Summary of Findings

Source of Variation	df.	Mean Square	F	p	R_p	p
Subtests	2	68.0919	53.8872	.001		
Students	35	0.7488	0.7488			
Subtests x Students	70	1.2636	1.2636		.791	.05
Residual	(182)	(1.00)				

TABLE I-31

ANALYSIS OF THEORETIC OVERACHIEVERS
PRUDENT-THEORETIC SCORES OF 0 AND 1

Subtest	Mean Stabilized Score	Grand Mean
2.00	8.7452	8.0607
3.00	8.8630	
4.00	6.5739	

Summary of Findings

Source of Variation	df.	Mean Square	F	p	R_P	p
Subtests	2	290.7335	276.4941	.001		
Students	174	0.9029	0.9029			
Subtests x Students	348	1.0515	1.0515		.951	.05
Residual	(182)	(1.00)				

TABLE I-32

ANALYSIS OF PRUDENT OVERACHIEVERS
PRUDENT-THEORETIC SCORES OF 2 AND 3

Subtest	Mean Stabilized Score	Grand Mean
2.00	8.4303	7.9276
3.00	8.8392	
4.00	6.5133	

Summary of Findings

Source of Variation	df.	Mean Square	F	p	R_P	p
Subtests	2	180.4188	147.9934	.001		
Students	116	0.7959	0.7959			
Subtests x Students	232	1.2191	1.2191		.820	.05
Residual	(182)	(1.00)				

TABLE I-33

ANALYSIS OF ALL OVERACHIEVERS HAVING A PRUDENT-THEORETIC SCORE

Subtest	Mean Stabilized Score	Grand Mean
2.00	8.6190	8.0073
3.00	8.8534	
4.00	6.5496	

Summary of Findings

Source of Variation	df.	Mean Square	F	p	R_p	p
Subtests	2	469.3902	418.7993	.001		
Students	291	0.8699	0.8699			
Subtests x Students	582	1.1208	1.1208		.892	.05
Residual	(182)	(1.00)				

TABLE I-34

ANALYSIS OF HIGHLY THEORETIC NORMAL ACHIEVERS

Subtest	Mean Stabilized Score	Grand Mean
2.00	7.2511	6.7060
3.00	7.3222	
4.00	5.5446	

Summary of Findings

Source of Variation	df.	Mean Square	F	p	R_p	p
Subtests	2	130.6626	92.4652	.001		
Students	128	1.3274	1.3274	.05		
Subtests x Students	256	1.4131	1.4131	.01		
Residual	(182)	(1.00)				

TABLE I-35

ANALYSIS OF HIGHLY THEORETIC UNDERACHIEVERS PROFILES ACROSS CATEGORIES 1.00, 3.00, AND 4.00

Subtest	Mean Stabilized Score	Grand Mean
1.00	6.3463	5.7219
3.00	6.1092	
4.00	4.7102	

Summary of Findings

Source of Variation	df.	Mean Square	F	p	R_p	p
Subtests	2	16.4175	9.8859	.001		
Students	20	0.5754	0.5754			
Subtests x Students	40	1.6607	1.6607	.025		
Residual	(182)	(1.00)				

TABLE I-36

ANALYSIS OF HIGHLY THEORETIC UNDERACHIEVERS PROFILES ACROSS FOUR TAXONOMY CATEGORIES

Subtest	Mean Stabilized Score	Grand Mean
1.00	6.3463	5.6751
2.00	5.5348	
3.00	6.1092	
4.00	4.7102	

Summary of Findings

Source of Variation	df.	Mean Square	F	p	R_p	p
Subtests	3	11.1289	7.6693	.001		
Students	20	0.5934	0.5934			
Subtests x Students	60	1.4511	1.4511	.05		
Residual	(182)	(1.00)				

TABLE I-37

ANALYSIS OF ALL NORMAL ACHIEVERS PROFILES ACROSS FOUR TAXONOMY CATEGORIES

Subtest	Mean Stabilized Score	Grand Mean
1.00	8.5578	7.0940
2.00	7.1274	
3.00	7.1893	
4.00	5.5016	

Summary of Findings

Source of Variation	df.	Mean Square	F	p	R_p	p
Subtests	3	2055.1066	1513.1104	.001		
Students	1315	1.2586	1.2586	.05		
Subtests x Students	3942	1.3582	1.3582	.01		
Residual	(182)	(1.00)				

APPENDIX J

EQUATING OF SCORES ON DIFFERENT EDITIONS OF SATO

APPENDIX J

EQUATING OF SCORES ON DIFFERENT EDITIONS OF SATO

SATO scores for students participating in the present study were obtained from four different editions of the test. The majority of students in this study wrote the 1963-64 edition of SATO.

The statistics for the different administrations of SATO were obtained by analyzing frequency distributions of the scores published by the Department of Educational Research of the Ontario College of Education, University of Toronto. A computer program[1] was obtained which generated various k statistics, from which Fisher's g statistics are derived. A modification to the program made by the present writer produced the g statistics and their standard errors. The results are presented in Tables J-1 and J-2.

TABLE J-1

SATO TOTAL VERBAL STATISTICS—GENERAL COURSE

Edition	N	Mean	SD	Skewness[a, b] g_1	Kurtosis[a, b] g_2
1961-62	37,148	33.4089	9.7588	.1306	-.5839
1962-63	38,250	34.2699	9.8163	.0828	-.6107
1963-64	44,029	26.4994	8.2948	.3661	-.2339
1964-65	48,410	33.4880	8.6439	.0989	-.4001

[a] Fisher's g_1 and g_2 statistics (Johnson, 1949, pp.153-158)
[b] All values significant at the .001 level

---

[1] UMKAY—written by Esther Schaeffer, University of Michigan, IBM SHARE Library No. 1548.

TABLE J-2

SATO MATHEMATICS STATISTICS—GENERAL COURSE

Edition	N	Mean	SD	Skewness[a, b] g_1	Kurtosis[a, b] g_2
1961-62	37,195	16.3708	5.8614	.0762	-.6571
1962-63	38,228	18.8270	5.4058	-.2771	-.4842
1963-64	44,016	17.1836	5.6865	-.0630	-.5958
1964-65	48,400	19.6575	5.5209	-.4010	-.4544

[a] Fisher's g_1 and g_2 statistics (Johnson, 1949, pp.153-158)

[b] All values significant at the .001 level

It is evident that the raw score distributions differ from year to year, particularly in skewness and kurtosis.

The comparison of scores from one edition of a test to another can be very misleading when raw scores only are used. Guilford (1956, Chap. 19) suggests that different score distributions be normalized and comparisons made of standard scores. He notes (p.491) that two conditions must be satisfied before accurate score comparisons can be made between different tests: (a) it must be assumed that the populations of students from which the distributions of scores arose have equal means and dispersions in ability from year to year, and (b) the form of each distribution, in terms of skewness and kurtosis, must be very similar from one test to another.

Condition (a) seems to have been met when one considers that over the years in which SATO was administered, no major educational change that affected Grade 12 General

Course students in Ontario took place; there seems to be no reason for assuming that that population's general aptitude would change markedly from year to year. Normalizing the distributions would bring about condition (b). Standardizing the scores would equalize the means and standard deviations of obtained scores.

Accordingly, a program written by the investigator was used to normalize and standardize all SATO scores. Scores thus obtained from the 1961-62, 1962-63, and 1964-65 distributions were then matched against the 1963-64 scores. For the purpose of the study, the raw scores of the 1961-62, 1962-63, and 1964-65 administrations were expressed in terms of equivalent 1963-64 raw scores, which were already entered on the majority of students' cards.

APPENDIX K

THE INVENTORY OF CHOICES - REPRODUCIBILITY, EQUIVALENCE, AND STABILITY OF SCALES; MIGRATION

APPENDIX K

THE INVENTORY OF CHOICES - REPRODUCIBILITY, EQUIVALENCE, AND STABILITY OF SCALES; MIGRATION

The need for adequate reproducibility of Guttman-type scales has been referred to in Chapter II. Since three different methods of scoring the Inventory of Choices were used, an indication of their agreement seems desirable. Attitude scales must possess a measure of stability over time if these scales are not to be of severely limited usefulness. In addition to the consistency of student scores as measured by test-retest correlation coefficients, migration of orientation, measured by changes in mean score for each attitude scale, should be considered.

Students in the writer's school were given the Inventory of Choices on two occasions seven months apart. The test-retest data and the data from the OTAC sample were analyzed by the Guttman scoring program at the University of California. Scores were based on the 12-point (item) scales created by Edwards and Wilson; in addition, 9-point scales were generated by eliminating the three least conforming items in each 12-point scale. As part of the analysis, correlation coefficients between the 9- and 12-point scales, as well as coefficients of reproducibility were computed.

Scoring for the Stouffer "H technique" 4-point scales was done by the present writer as described in Chapter IV. Coefficients of reproducibility were calculated by hand, using the procedure described by Edwards and Wilson (1958a, pp.34-35); correlation coefficients were generated by a program written by the investigator. In every case the correlation coefficient calculated was the Pearson product-moment coefficient. The suitability of this statistic for the data available is pointed out by Kendall and Stuart (1961, pp.566-567), who illustrate the technique for its computation when discrete data are used as input.

To measure migration on the 4-point Inventory of Choices scales the t test for significance of changes in correlated means was used (Popham, 1967, p.152); for the 9- and 12-point scales, t tests were obtained as part of the processing carried out at the University of California.

Reproducibility

Coefficients of reproducibility are presented in Table K-1.

TABLE K-1

COEFFICIENTS OF REPRODUCIBILITY

	OTAC Population N = 2250			Reported by Edwards and Wilson[a]
	Scale			Scale
	4-point	9-point	12-point	4-point
Prudent-Theoretic	.932	.893	.867	.94
Prudent-Immediate	.872	.877	.852	.87
Prudent-Aesthetic	.916	.878	.850	.89
Theoretic-Immediate	.894	.883	.858	.91
Theoretic-Aesthetic	.893	.872	.857	.91
Aesthetic-Immediate	.877	.871	.857	.89

[a] Edwards and Wilson (1958a, pp.35-40).

It will be noted that the coefficients of reproducibility are somewhat lower for the 9- and 12-point scales than for the 4-point scales, whose reproducibilities are about the same as those reported by Edwards and Wilson.

Equivalence

Table K-2 presents the results of correlating 4-, 9-, and 12-point scales with each other within each scale of the Inventory.

TABLE K-2

EQUIVALENCE OF THREE SCORING METHODS

N = 2250

	Scale Pairs		
	4- and 9-point	4- and 12-point	9- and 12-point
Prudent-Theoretic	.5764	.6208	.6789
Prudent-Immediate	.6485	.5702	.5769
Prudent-Aesthetic	.5609	.5539	.4154
Theoretic-Immediate	.5463	.5709	.5476
Theoretic-Aesthetic	.5801	.5912	.5642
Aesthetic-Immediate	.6101	.5910	.6371

The correlations in Table K-2 indicate that the three scaling methods do not agree highly with each other.

Stability

Maykovich (1966) showed that long-term constancy of Inventory of Choices scores could not be assumed. A question not answered by Maykovich's study is whether these scores remain relatively stable over the Grade 12 year.

Table K-3 presents the results of correlating the scores of the first administration with those of the second administration given seven months later.

TABLE K-3

STABILITY (TEST-RETEST CORRELATION) COEFFICIENTS
N = 155

	Scale		
	4-point	9-point	12-point
Prudent-Theoretic	.6126	.4968	.5100
Prudent-Immediate	.4866	.3228	.2170
Prudent-Aesthetic	.3719	.4366	.3377
Theoretic-Immediate	.4590	.5041	.4842
Theoretic-Aesthetic	.4709	.3964	.3377
Aesthetic-Immediate	.5422	.4300	.4513

Correlation coefficients in Table K-3 appear to be quite low compared to typical test-retest coefficients of achievement tests. It must be kept in mind that attitudes are being measured and that measurements of these must be expected to show less reliability than do tests made of achievement items; for example, Cronbach (1960, p.140) cites the stability coefficients of the Allport-Vernon Study of Values as ranging from .39 to .84 for six scores retested after three months.

Migration

While stability gauges whether students maintain their relative position on each score continuum over a period of time, migration measures changes in the group mean of each score continuum after a period of time. Table K-4 lists the scales in which significant migration was observed over the seven month interval between testings; scales which are not listed did not show significant changes in means.

TABLE K-4

MIGRATION OF ORIENTATION DURING THE GRADE 12 YEAR

Scale	Pretest Mean	Posttest Mean	Significance level of t	Migration away from
Prudent-Theoretic 4-point	2.577	2.729	.05	Theoretic
Prudent-Aesthetic 4-point	2.296	2.118	.05	Prudent
Theoretic-Immediate 4-point	2.512	2.269	.01	Theoretic
Theoretic-Aesthetic 4-point	2.415	2.098	.01	Theoretic
Theoretic-Immediate 9-point	4.389	5.088	.01	Theoretic
Theoretic-Aesthetic 9-point	4.000	4.626	.05	Theoretic
Theoretic-Immediate 12-point	5.897	6.637	.05	Theoretic
Theoretic-Aesthetic 12-point	5.074	5.690	.05	Theoretic

Note:

The Stouffer (4-point) scores are opposite in direction to the Guttman (9- and 12-point) scores; for example a high score on the Theoretic-Immediate continuum indicates a Theoretic orientation on the Stouffer scale, but an Immediate orientation on the Guttman scales.

Examination of Table K-4 reveals significant migration away from the Theoretic pole in all three 4-point scales containing the Theoretic orientation as an alternative, and a similar migration on two of these three scales scored on a 9- or 12-point basis. While equivalence coefficients show that the three methods of scaling do not agree highly with one another, it is interesting to note that all three methods of scoring agree to the extent observed in detecting migration of orientation.

Summary

The Inventory of Choices scales are seen to have adequate reproducibility and moderate stability; moderate equivalence exists amongst scales scored by three different methods. While considerable fluctuation in individual scores was observed over a seven month interval, during that time a significant change in mean score, indicative of migration away from the Theoretic orientation, was generally observed for the group studied.

It was observed in Chapter V that the 9- and 12-point scoring systems for the Theoretic-Immediate and Prudent-Theoretic scales were no more effective in reducing unexplained variance than was the corresponding 4-point scale; these results and the results reported here favor the use of the 4-point scales over the 9- or 12-point scales.[1] The results of the analyses of covariance reported in Appendix L lead one to believe that a 4-point scale itself is too fine a measuring device to be used in an instrument such as the Inventory of Choices. The results of Chapter V likewise indicate that extreme positions on the scale are significant and that a 2- or 3-point scale might more adequately reflect the instrument's resolving power. The wisdom of Maykovich's concentration on extreme positions on a scale is borne out by the findings of this study.

---

[1] An advantage would be gained if the responses were to be scored by computer: a program for the Stouffer H-technique of scoring the Inventory of Choices would be quite simple to write and inexpensive to use, compared to the intricate Guttman scoring program.

APPENDIX L

ANALYSES OF COVARIANCE

APPENDIX L

ANALYSES OF COVARIANCE

In this Appendix total test scores and subtest scores are considered independently of one another. In addition to OTAC scores, teacher-assigned final chemistry marks and final average marks are analyzed.[1] The analyses are presented to facilitate comparison of this study with the studies of Edwards and Wilson, Maykovich, and others who have investigated the relationships of non-intellective factors to achievement in high school.

The analyses presented herewith have the advantage that a double covariance analysis provides, namely, that SATO Total Verbal and SATO Mathematics scores are accounted for separately in adjusting the means of the treatment groups.

The writer had access to the multiple analysis of covariance program developed by Dr. L.D. McLean and Mr. Paul Barbuto of The Ontario Institute for Studies in Education. The significance of differences between pairs of means was tested by the method of Scheffé[2] in a program written by the investigator.[3]

---

[1] It will be recalled that final chemistry marks were treated to remove the OTAC component, and that final average marks do not include the final chemistry marks (pp.125-126).

[2] Guenther, William C., Analysis of Variance (Englewood Cliffs, N.J.; Prentice-Hall Inc., 1964) pp.57-58, 149-150.

[3] The writer is indebted to Dr. R.P. Bhargava of OISE for deriving the variance estimate of the difference between adjusted means when two covariates are used.

Since the following analyses were not intended to form the major part of this investigation, but were intended merely to illuminate other findings, condensed results are reported in the following tables.

TABLE L-1

ASSOCIATION BETWEEN THEORETIC-IMMEDIATE SCORES AND OTAC MEANS ADJUSTED FOR SATO TOTAL VERBAL AND MATHEMATICS SCORES

Criterion	Theoretic-Immediate Score 0 N=324	 1 N=706	 2 N=602	 3 N=208	F	p
Total OTAC	23.68	24.84	* 25.96	27.64	21.30	.01
Category 1.00	9.12	9.53	* 10.06	10.71	14.94	.01
Category 2.00	4.95	5.27	5.32	5.72	7.10	.01
Category 3.00	5.67	6.09	* 6.45	6.83	17.40	.01
Category 4.00	3.95	3.95	4.14	4.38	4.26	.01

*Difference between two adjacent mean scores significant at the .05 level.

TABLE L-2

ASSOCIATION BETWEEN PRUDENT-THEORETIC SCORES AND OTAC MEANS ADJUSTED FOR SATO TOTAL VERBAL AND MATHEMATICS SCORES

Criterion	Prudent-Theoretic Score 0 N=282	 1 N=622	 2 N=640	 3 N=403	F	p
Total OTAC	28.34	** 25.74	* 24.48	23.71	36.09	.01
Category 1.00	11.05	** 9.91	9.45	9.20	24.21	.01
Category 2.00	6.01	** 5.37	5.05	5.00	19.33	.01
Category 3.00	6.84	6.40	6.04	5.77	17.16	.01
Category 4.00	4.44	4.06	3.94	3.74	8.86	.01

*Difference between two adjacent mean scores significant at the .05 level.

**Difference between two adjacent mean scores significant at the .01 level.

TABLE L-3

ASSOCIATION BETWEEN THEORETIC-IMMEDIATE SCORES AND TEACHER-ASSIGNED MARKS (GRADES). MEANS ADJUSTED FOR SATO TOTAL VERBAL AND MATHEMATICS SCORES

Criterion	Theoretic-Immediate Score				F	p
	0 N=324	1 N=706	2 N=602	3 N=208		
Final Chemistry Mark	53.82	** 59.55	** 62.99	64.34	33.00	.01
Final Average Mark	60.34	61.17	61.82	63.61	3.47	.05

**Difference between two adjacent mean scores significant at the .01 level.

TABLE L-4

ASSOCIATION BETWEEN PRUDENT-THEORETIC SCORES AND TEACHER-ASSIGNED MARKS (GRADES). MEANS ADJUSTED FOR SATO TOTAL VERBAL AND MATHEMATICS SCORES

Criterion	Prudent-Theoretic Score				F	p
	0 N=282	1 N=622	2 N=640	3 N=403		
Final Chemistry Mark	64.24	62.19	** 58.46	57.56	17.06	.01
Final Average Mark	62.46	61.65	61.07	61.70	0.91	-

**Difference between two adjacent mean scores significant at the .01 level.

Examination of Tables L-1 to L-4 reveals that the F's for differences in means among treatment groups are significant at the .01 level in all but two cases. Although the trends of increasing or decreasing means are quite evident, the differences between adjacent means generally are not significant. These observations indicate that while the two Inventory of Choices scales are functioning, 4-point scales are too fine to use with the criterion measures.

The significant F's suggest that a dichotomized scale might be more useful in making significant discriminations. Since there are three ways of dichotomizing each scale while maintaining the original scale order, selection of a suitable dichotomizing point is of some concern. The choice of a convenient dividing point is empirically investigated in the tables which follow.

TABLE L-5

ASSOCIATION BETWEEN DICHOTOMIZED THEORETIC-IMMEDIATE SCORES AND OTAC MEANS ADJUSTED FOR SATO TOTAL VERBAL AND MATHEMATICS SCORES

Criterion	Theoretic-Immediate Score		F	p
	0 N = 324	1,2, & 3 N = 1516		
Total OTAC	23.68	25.67	27.67	.01
Category 1.00	9.12	9.90	17.59	.01
Category 2.00	4.95	5.35	12.01	.01
Category 3.00	5.67	6.34	28.15	.01
Category 4.00	3.95	4.08	1.62	-
	0 & 1 N = 1030	2 & 3 N = 810		
Total OTAC	24.48	26.39	43.70	.01
Category 1.00	9.40	10.22	33.23	.01
Category 2.00	5.17	5.42	7.83	.01
Category 3.00	5.96	6.55	36.95	.01
Category 4.00	3.95	4.20	9.60	.01
	0,1, & 2 N = 1632	3 N = 208		
Total OTAC	25.03	27.63	32.88	.01
Category 1.00	9.64	10.71	22.52	.01
Category 2.00	5.23	5.72	12.47	.01
Category 3.00	6.14	6.83	20.21	.01
Category 4.00	4.02	4.38	8.17	.01

TABLE L-6

ASSOCIATION BETWEEN DICHOTOMIZED PRUDENT-THEORETIC SCORES AND OTAC MEANS ADJUSTED FOR SATO TOTAL VERBAL AND MATHEMATICS SCORES

Criterion	Prudent-Theoretic Score		F	p
	0 N = 282	1,2, & 3 N = 1665		
Total OTAC	28.29	24.77	78.37	.01
Category 1.00	11.03	9.56	57.25	.01
Category 2.00	6.00	5.16	45.77	.01
Category 3.00	6.82	6.11	27.58	.01
Category 4.00	4.43	3.94	18.72	.01
	0 & 1 N = 904	2 & 3 N = 1043		
Total OTAC	26.53	24.20	67.18	.01
Category 1.00	10.26	9.36	41.63	.01
Category 2.00	5.57	5.04	35.39	.01
Category 3.00	6.53	5.94	38.32	.01
Category 4.00	4.17	3.87	14.10	.01
	0,1, & 2 N = 1544	3 N = 403		
Total OTAC	25.68	23.77	29.68	.01
Category 1.00	9.92	9.22	16.72	.01
Category 2.00	5.35	5.02	9.18	.01
Category 3.00	6.33	5.78	21.56	.01
Category 4.00	4.08	3.75	10.95	.01

TABLE L-7

ASSOCIATION BETWEEN DICHOTOMIZED THEORETIC-IMMEDIATE SCORES AND TEACHER-ASSIGNED MARKS (GRADES). MEANS ADJUSTED FOR SATO TOTAL VERBAL AND MATHEMATICS SCORES

Criterion	Theoretic-Immediate Score		F	p
	0 N = 324	1,2, & 3 N = 1516		
Final Chemistry Mark	58.83	61.57	63.08	.01
Final Average Mark	60.34	61.76	3.74	-
	0 & 1 N = 1030	2 & 3 N = 810		
Final Chemistry Mark	57.75	63.33	63.08	.01
Final Average Mark	60.91	62.28	5.88	.05
	0,1, & 2 N = 1632	3 N = 208		
Final Chemistry Mark	59.68	64.31	17.23	.01
Final Average Mark	61.25	63.60	7.11	.01

TABLE L-8

ASSOCIATION BETWEEN DICHOTOMIZED PRUDENT-THEORETIC SCORES AND TEACHER-ASSIGNED MARKS (GRADES). MEANS ADJUSTED FOR SATO TOTAL VERBAL AND MATHEMATICS SCORES

Criterion	Prudent-Theoretic Score		F	p
	0 N = 282	1,2, & 3 N = 1665		
Final Chemistry Mark	64.11	59.66	21.22	.01
	0 & 1 N = 904	2 & 3 N = 1043		
Final Chemistry Mark	62.81	58.13	46.40	.01
	0,1, & 2 N = 1544	3 N = 403		
Final Chemistry Mark	60.98	57.68	15.37	.01

Note:

The means of Final Average Marks are not included since in Table L-4 no significant differences in those means were observed.

From examination of Tables L-5 to L-8 it is evident that dichotomizing either scale produces significant differences in means, and does so, in nearly all cases, regardless of the method of dichotomization. It would seem then that a 2-point scale more accurately reflects the resolving power of the Theoretic-Immediate and Prudent-Theoretic scales, at least with the criterion measures available in this study.

Since the use of a dichotomized scale prevents one from detecting curvilinear relationships, the construction of suitable 3-point scales might profitably be attempted for the Inventory of Choices.

APPENDIX M

COMPARISON OF TEACHER-ASSIGNED MARKS (GRADES) AND OBJECTIVE TEST SCORES

APPENDIX M

COMPARISON OF TEACHER-ASSIGNED MARKS (GRADES) AND OBJECTIVE TEST SCORES

Final chemistry marks and final average marks were collected for most students participating in this study. The relationships of these marks to OTAC and SATO scores, the effect of selected independent variables in explaining the variance of OTAC scores and final chemistry mark, and the results of analyses of covariance on OTAC Total scores and teacher-assigned marks are considered here.

Table M-1 presents the relationships found to exist between teacher-assigned marks and OTAC scores.

TABLE M-1

CORRELATIONS BETWEEN FINAL CHEMISTRY MARK, FINAL AVERAGE MARK, AND OTAC SCORES

Score	Final Chemistry Mark[a]	Final Average Mark[b]
OTAC Total	.5634	.4527
Category 1.00	.4841	.3624
Category 2.00	.4387	.3579
Category 3.00	.4907	.4087
Category 4.00	.3170	.2817
Final Chemistry Mark	-	.6954

[a] Treated to remove OTAC component.

[b] Chemistry mark not included in average.

Examination of Table M-1 shows that teacher-assigned marks are not highly related to OTAC scores. As would be expected, the relationships between OTAC scores and final chemistry marks are higher than the relationships between OTAC scores and final average marks. It is interesting to note that the relationship between final chemistry marks and final average marks (which do not include chemistry marks) is higher than any other relationship in the table. This observation indicates that marks assigned by teachers of chemistry agree more closely with marks assigned by the teachers of other subjects than with OTAC scores. It is evident that OTAC and teachers of chemistry are to a large extent not measuring the same accomplishments.

The correlations of Category 2.00 are somewhat lower than those of Categories 1.00 and 3.00, and the correlations of Category 4.00 are considerably lower than the others. Assuming that the Category 4.00 subtest measures what it purports to measure, its low relationship with grades suggests that comparatively little emphasis is placed, either in teaching or in examining, on the achievement of abilities which are subsumed under the cognitive objective Analysis.

Table M-2 presents the correlations of teacher-assigned grades with SATO scores. For ease of comparison the correlations of OTAC with SATO scores (Table 21) are repeated.

TABLE M-2

CORRELATIONS OF SATO SCORES WITH TEACHER-ASSIGNED MARKS AND OTAC SCORES

Mark or Score	SATO Total Verbal	SATO Mathematics
Final Chemistry Mark[a]	.3235	.3704
Final Average Mark[b]	.4081	.3245
OTAC Total	.5306	.5834
Category 1.00	.4393	.4339
Category 2.00	.4331	.5255
Category 3.00	.4506	.5654
Category 4.00	.3236	.2895
SATO Total Verbal	-	.4907

[a] Treated to remove OTAC component.

[b] Chemistry mark not included in average.

Inspection of Table M-2 shows that teacher-assigned marks are not as highly related to scholastic aptitude as are most OTAC scores, with the exception of Category 4.00 scores.

The correlation of final chemistry marks to SATO Mathematics is higher than the correlation to SATO Total Verbal; a similar relationship is observed for OTAC Total scores, Category 2.00 scores, and Category 3.00 scores. The reverse relationship is observed with final average marks and SATO scores; since, for most students, final average marks will include a mark in mathematics, this reversal could be expected to be more pronounced were it possible to remove the mathematics mark from the computation of the final average mark.

Considering the results of Tables M-1 and M-2 together, it is seen that the relationship of OTAC Total scores to final chemistry marks is approximately equal to the relationships of OTAC Total scores to SATO Total Verbal or SATO Mathematics. For OTAC subtests the correlations to final chemistry marks are approximately the same as to SATO Total Verbal scores, and generally less than the subtests' correlations to SATO Mathematics scores.

The first educational hypothesis (pp.109-110) is supported in the case of OTAC Total scores; for OTAC subtest scores the relationships are higher than was expected.

The Automatic Interaction Detector program yields comparisons of OTAC scores and final chemistry marks which are not evident in correlation studies. Run No. 1 (Figure 4) and run No. 5 (Figure 8) both used the same set of explanatory variables. Comparison of the two AID trees yields the following observations.

SATO Mathematics is the best splitting variable for both OTAC scores and final chemistry marks; both splits are very much alike.

SATO Total Verbal is the next best splitter, for both OTAC scores and final chemistry marks, for those students obtaining an above-average score in SATO Mathematics.

For students who do not obtain an above-average score in SATO Mathematics, SATO Total Verbal, the next best splitter for OTAC scores, is not as effective as immediate educational plans in splitting groups where final chemistry marks is the dependent variable. In other branches and twigs of the final chemistry mark AID tree, immediate educational plans supplants SATO Total Verbal as a predictor.

The Theoretic-Immediate variable operates most effectively in a different verbal aptitude range in explaining final chemistry marks than it does in explaining OTAC scores.

The Prudent-Theoretic variable does not effectively split any final chemistry marks group.

The variable Repeating, which did not function effectively in explaining OTAC Total scores, is a good predictor of final chemistry mark for students of moderate to low mathematical ability who do not plan to enter university.

Table M-3 compares the percentages of variance explained by effective splitters for the two dependent variables.

TABLE M-3

CONTRIBUTION OF IMPORTANT VARIABLES TO OTAC TOTAL SCORE VARIANCE AND FINAL CHEMISTRY MARK VARIANCE

Variable	Percent of Variance Explained	
	for OTAC Total Score (Run No.1)	for Final Chemistry Mark
SATO Math	27.55	11.62
SATO TV	10.66	1.97
Ed Plans 1	1.50	6.52
T-I 4	1.02	0.50
P-T 4	0.54	-
Repeating	-	0.90
Total	41.27	21.51

It is seen that those independent variables which are useful as predictors for both dependent variables operate to explain the variance of the dependent variable in quite different ways. The ratios of SATO Mathematics to SATO Total Verbal in terms of variance explained is especially striking. Also of note is the effectiveness of immediate educational plans compared to SATO Total Verbal as a predictor of final chemistry marks.

The interactions SATO Mathematics x SATO Total Verbal and immediate educational plans x SATO Total Verbal are common to both AID analyses. For final chemistry marks the three other interactions each contain immediate educational plans as a component; this variable is not predominant in the interactions observed in the OTAC Total score AID tree.

The analyses of covariance reported in Appendix L show that when students are grouped according to their Theoretic-Immediate scores or Prudent-Theoretic scores and their means adjusted for SATO Total Verbal and SATO Mathematics scores, Theoretic students have higher final chemistry marks than those of other orientations; Theoretic students also have higher OTAC Total scores and subtest scores. On the other hand the relationship of final average marks to Theoretic-Immediate scores was not as pronounced as that observed with final chemistry marks. Prudent-Theoretic groupings had no significant effect on final average marks.

APPENDIX N

EQUATIONS FOR COMPUTING RESIDUAL SCORES AND THEIR STANDARD ERRORS OF MEASUREMENT

APPENDIX N

EQUATIONS FOR COMPUTING RESIDUAL SCORES AND THEIR STANDARD ERRORS OF MEASUREMENT

Regression Equations

A special form of the regression equation for three variables is given by Garrett (1954, pp.391-392):

$$\hat{x}_1 = \frac{\sigma_1(r_{12} - r_{13}r_{23})}{\sigma_2(1 - r_{23}^2)} \cdot x_2 + \frac{\sigma_1(r_{13} - r_{12}r_{23})}{\sigma_3(1 - r_{23}^2)} \cdot x_3 \quad (1)$$

where $\hat{x}_1$ is the predicted deviation score on the dependent variable

x_2 is the deviation score obtained on the first predictor

x_3 is the deviation score obtained on the second predictor

σ_1 is the standard deviation of the dependent variable

σ_2 is the standard deviation of the first predictor

σ_3 is the standard deviation of the second predictor

r_{12} is the correlation between the dependent variable and first predictor

r_{13} is the correlation between the dependent variable and the second predictor

r_{23} is the correlation between the first and second predictors.

The coefficients of x_2 and x_3 may be replaced by b_2 and b_3 respectively.

Thus

$$b_2 = \frac{\sigma_1(r_{12} - r_{13}r_{23})}{\sigma_2(1 - r_{23}^2)} \tag{2}$$

$$b_3 = \frac{\sigma_1(r_{13} - r_{12}r_{23})}{\sigma_3(1 - r_{23}^2)} \tag{3}$$

where b_2 is the partial regression coefficient for x_2

and b_3 is the partial regression coefficient for x_3.

Replacing deviation scores with raw scores and using the partial regression coefficients b_2 and b_3 we may rewrite equation (1) as follows:

$$(\hat{X}_1 - M_1) = b_2(X_2 - M_2) + b_3(X_3 - M_3) \tag{4}$$

where $\hat{X}_1$ is the predicted raw score on the dependent variable

X_2 is the observed raw score on the first predictor

X_3 is the observed raw score on the second predictor

and M_1, M_2 and M_3 are the means of the scores on the dependent and first and second predictor variables respectively.

Solving for $\hat{X}_1$ we obtain

$$\hat{X}_1 = b_2X_2 + b_3X_3 - b_2M_2 - b_3M_3 + M_1 \tag{5}$$

The last three terms are constants and may be replaced with their sum C. Thus the predicted score on the dependent variable is found by

$$\hat{X}_1 = b_2X_2 + b_3X_3 + C \tag{6}$$

The residual score R is simply

$$R = X_1 - \hat{X}_1 \tag{7}$$

where X_1 is the obtained score

and $\hat{X}_1$ is the score predicted by equation (6).

Adjustment of Standard Error of Measurement

Let subscript 1 denote OTAC pretest

2 denote OTAC posttest

3 denote SATO Verbal

4 denote SATO Mathematics.

The partial correlation between pretest and posttest, with SATO Verbal held constant, is given (Guilford, 1956, p.316-317) by

$$r_{12.3} = \frac{r_{12} - r_{13} \cdot r_{23}}{\sqrt{(1 - r_{13}^2)(1 - r_{23}^2)}}$$

Similarly

$$r_{14.3} = \frac{r_{14} - r_{13} \cdot r_{34}}{\sqrt{(1 - r_{13}^2)(1 - r_{34}^2)}}$$

and

$$r_{24.3} = \frac{r_{24} - r_{23} \cdot r_{24}}{\sqrt{(1 - r_{23}^2)(1 - r_{34}^2)}}$$

The partial correlation between pretest and posttest with both SATO Total Verbal and SATO Mathematics held constant is then given by

$$r_{12.34} = \frac{r_{12.3} - r_{14.3} \cdot r_{24.3}}{\sqrt{(1 - r_{14.3}^2)(1 - r_{24.3}^2)}}$$

The variabilities of the pretest and posttest freed from the influence of SATO Total Verbal and SATO Mathematics is given (Garrett, 1954, p.390) by

$$\sigma_{1.34} = \sigma_1 \sqrt{1 - r_{13}^2} \cdot \sqrt{1 - r_{14.3}^2}$$

$$\sigma_{2.34} = \sigma_2 \sqrt{1 - r_{23}^2} \cdot \sqrt{1 - r_{24.3}^2}$$

Using the procedure followed by Haggard in analyzing Rosen's data (Haggard, 1958, Table 11, p.114; Rosen, 1953) the combined variability estimate is given by

$$\sigma_{\overline{12}.34} = \sqrt{\frac{\sigma_{1.34}^2 + \sigma_{2.34}^2}{2}}$$

Thus the adjusted standard error of measurement, in raw score units, is

$$s.e._{meas} = \sigma_{\overline{12}.34} \sqrt{1 - r_{12.34}^2}$$

Now Haggard's method of pattern analysis requires that the scores be standardized; in this study residual T scores with a standard deviation of 10 were used and thus

$$\frac{\sigma_{T.34}}{\sigma_{\overline{12}.34}} = \frac{10}{\sigma_{big.34}}$$

from which

$$\sigma_{T.34} = \frac{10\sigma_{\overline{12}.34}}{\sigma_{big.34}}$$

where the subscript T.34 denotes residuals placed on a T-scale basis and "big" denotes the main OTAC distribution. $\sigma_{big.34}$ is obtained empirically from the frequency distribution of OTAC residuals derived from the main administration of OTAC. The adjusted error of measurement in residual T-score units is therefore

$$s.e._{meas.(T)} = \sigma_{T.34}\sqrt{1 - r_{12.34}}$$

This formula represents the standard error of measurement free of the influence of SATO Total Verbal and SATO Mathematics scores and is used to compute the values entered in the last column of Table 22.

BIBLIOGRAPHY

BIBLIOGRAPHY

Ahmann, J. Stanley, and Glock, Marvin D. Evaluating Pupil Growth. 2d ed. Boston: Allyn and Bacon, Inc., 1963.

Allen, Hugh, Jr. Attitudes of Certain High School Seniors Toward Science and Scientific Careers. Science Manpower Project Monographs. New York: Bureau of Publications, Teachers College, Columbia University, 1959.

Analyses of Science Tests. Washington: National Science Teachers Association, 1959.

Anderson Chemistry Test. Forms AM and BM, 1951, 1952. Tarrytown, N.Y.: Harcourt, Brace and World.

Anderson, June Sutton. "A Comparative Study of Chemical Educational Material Study and Traditional Chemistry in Terms of Students' Ability to Use Selected Cognitive Processes." Unpublished Ph.D. dissertation, The Florida State University, Tallahassee, 1964. Dissertation Abstracts (1965), 5147-5148.

Anderson, Kenneth E. "A Frontal Attack on the Basic Problem in Evaluation: The Achievement of the Objectives of Instruction in Specific Areas." Journal of Experimental Education, 18 (1950), 163-174.

________. "The Relative Achievements of the Objectives of Secondary School Science in a Representative Sampling of Fifty-six Minnesota Schools." Unpublished Ph.D. dissertation, The University of Minnesota, 1949.

Arkin, Herbert, and Colton, Raymond R. Tables for Statisticians. New York: Barnes & Noble, Inc., 1953.

Ausubel, David P. "An Evaluation of the Conceptual Schemes Approach to Science Curriculum Development." Journal of Research in Science Teaching, 3 (1965), 255-264.

Ayers, J. Douglas. "Justification of Bloom's Taxonomy by Factor Analysis." Paper read at the American Educational Research Association, Chicago, February, 1966.

_________. et al. Summary Description of Grade Nine Science Objectives and Test Items. Edmonton, Alberta: The High School Entrance Examinations Board, Department of Education, March, 1965.

Aylesworth, Thomas G. "Problem-Solving: A Comparison of the Expressed Attitudes with the Classroom Methodology of Science Teachers in Selected High Schools." Science Education, 44 (1960), 366-374.

Bebel, Clifford. "Evaluation and Curriculum Development." Educational Leadership, 20 (October, 1962), 4-6.

Beauchamp, Wilbur L. Instruction in Science. U.S. Office of Education, Bulletin 1932, No. 17, Monograph No. 22. Washington, D.C.: U.S. Government Printing Office, 1932.

Behnke, Frances L. "Opinions of a Selected Group of High School Science Teachers and Scientists on Some Issues Related to Science and Science Teaching." Unpublished Doctoral Thesis, Teachers College, Columbia University, New York City, 1959.

Blanc, Sam S. "Review of the General Goals in Science Teaching." Science Education, 36 (1952), 47-52.

Bloom, Benjamin S., et al. Taxonomy of Educational Objectives. Handbook I: Cognitive Domain. New York: Longmans, Green & Co., 1956.

Bronowski, Jacob. "Science as Foresight." What is Science? Edited by J. R. Newman. New York: Simon & Schuster, 1955.

_________. The Common Sense of Science. London: William Heinemann, 1951.

Brown, M. G. "New Thinking in School Chemistry." Chemistry and Industry, No. 13 (March 31, 1962), 591-597.

Burnett, R. Will. Teaching Science in the Secondary Schools. New York: Rinehart and Co., Inc., 1957.

Butts, David P. "The Evaluation of Problem Solving in Science." Journal of Research in Science Teaching. 2 (1964), 116-122.

California Test of Mental Maturity. [Grades 4-6], Monterey, Calif.: California Test Bureau.

Carleton, Robert H., et al. "Improving Secondary School Science." Rethinking Science Education. The Fifty-Ninth Yearbook of the National Society for the Study of Education, Part I. Edited by Nelson B. Henry. Chicago: The University of Chicago Press, 1960.

Charen, George. "The Effect of Open-Ended Experiments in Chemistry on the Achievement of Certain Objectives in Science Teaching." Journal of Research in Science Teaching, 1 (1963), 184-190.

Conant, James B. On Understanding Science. New Haven, Connecticut: Yale University Press, 1947.

________. "Scientific Education of the Layman." Yale Review, 36 (September, 1946), 15-36.

Cook, Desmond L. "A Note on Relevance Categories and Item Statistics." Educational and Psychological Measurement, 20 (1960), 321-331.

Cooley, William W., and Klopfer, Leopold E. "The Evaluation of Specific Educational Innovations." Journal of Research in Science Teaching, 1 (1963), 73-80.

Cooperative Science Tests Handbook. Princeton, N.J.: Educational Testing Service, 1964.

Cornell Class-Reasoning Test. Form X, (1964). Ithaca, N.Y.: Cornell Critical Thinking Project, Stone Hall.

Cornell Conditional-Reasoning Test. Form X, (1964). Ithaca, N.Y.: Cornell Critical Thinking Project, Stone Hall.

Cox, Richard C. "In-Progress Studies and Utilization of the Taxonomy." University of Pittsburg School of Education, 1966. (Mimeographed.)

________. "Item Selection Techniques and Evaluation of Instructional Objectives." Journal of Educational Measurement, 2 (1965), 181-185.

Cox, Richard C., and Gordon, John M. "In-Progress Studies and Utilization of the *Taxonomy*." University of Pittsburgh School of Education, 1966. (Mimeographed.)

Cox, Richard C., and Unks, Nancy Jordan. *A Selected and Annotated Bibliography of Studies Concerning the Taxonomy of Educational Objectives*: *Cognitive Domain*. Pittsburgh, Pa.: Learning Research and Development Center, University of Pittsburgh, 1967.

Cragg, L. H., Graham, R. P., and Young, J. V. *The Elements of Chemistry*. Toronto: Clarke, Irwin & Co. Ltd., 1959.

Croal, A. G., Couke, J. H., and Louden, A. H. *Chemistry for Secondary Schools*. Rev. ed. Toronto: The Copp Clark Publishing Co. Ltd., 1958.

Cronbach, Lee J. *Essentials of Psychological Testing*. 2d ed. New York: Harper and Brothers, 1960.

Davis, Frederick, B. *Educational Measurements and Their Interpretation*. Belmont, Calif.: Wadsworth Publishing Co., Inc., 1964.

Dixon, Wilfrid J. and Massey, Frank J., Jr. *Introduction to Statistical Analysis*. New York: McGraw-Hill Book Company, Inc., 1957.

Dressel, Paul L., *et al*. "How the Individual Learns Science." *Rethinking Science Education*. The Fifty-Ninth Yearbook of the National Society for the Study of Education, Part I. Edited by Nelson B. Henry. Chicago: The University of Chicago Press, 1960.

________, and Mayhew, Lewis B. *General Education*: Explorations in Evaluation. Washington, D.C.: American Council on Education, 1954.

________, and Nelson, Clarence. "Questions and Problems in Science." Test Folio No. 1. Princeton, N.J.: Cooperative Test Division, Educational Testing Service, 1956.

Dyer, Henry S. "On the Assessment of Academic Achievement." *Teachers College Record*, *62* (November, 1960), 164-172.

Ebel, Robert L. "How an Examination Service Helps College Teachers to give Better Tests." *Proceedings of the Invitational Conference on Testing Problems*. Princeton, N.J.: Educational Testing Service, 1953.

Ebel, Robert L. *Measuring Educational Achievement*. Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1965.

Edwards, T. Bentley, and Wilson, Alan B. "The Association Between Interest and Achievement in High School Chemistry." *Educational and Psychological Measurement* (Winter, 1959), 601-610. (b)

_________. *Attitudes of High School Students as Related to Success in School*. Final Report on Project SAE 7143 "Attitudes of High School Students as Related to Success in School. Berkeley, Calif.: Department of Education, University of California, July 31, 1958. (a)

_________. "Attitudes Towards the Study of School Subjects." *Educational Theory*, *8* (October, 1958), 275-284. (c)

_________. "The Development of Scales of Attitudinal Dimensions." *Journal of Experimental Education*, *28* (September, 1959), 3-36. (a)

_________. "The Specialization of Interests and Academic Achievement." *Harvard Educational Review*, *27* (Summer, 1958), 183-196. (b)

_________. *A Study of Some Social and Psychological Factors Influencing Educational Achievement*. Final Report on Project SAE 7787 "The Specialization of Attitudes as Related to Academic Success at Varying Academic Levels." Berkeley, Calif.: Department of Education, University of California, June 30, 1961.

Encyclopedia of Educational Research. 3rd. ed. 1960.

Examinations Bulletin No. 3. The Certificate of Secondary Education: An Introduction to Some Techniques of Examining. Secondary School Examinations Council, 1964. Her Majesty's Stationery Office: London, England.

Examinations Bulletin No. 8. The Certificate of Secondary Education: Experimental Examinations: Science. The Schools Council, 1965. Her Majesty's Stationery Office: London, England.

Ferguson, George A. *Statistical Analysis in Psychology and Education*. 2d ed. McGraw-Hill Book Co., Inc., 1966.

Fischler, Abraham S. "Challenge of Science Teaching Today and Tomorrow." *Science Education*, *47* (1963), 348-353.

Fischler, Abraham S. "Science, Process, The Learner: A Synthesis." Science Education, 49 (1965), 402-409.

Fitzpatrick, Frederick L. Policies for Science Education. Science Manpower Project Monographs, New York: Bureau of Publications, Teachers College, Columbia University, 1960.

Furst, Edward J. "The Effect of the Organization of Learning Experiences upon the Organization of Learning Outcomes." Journal of Experimental Education, 18 (1950), 215-228, 343-352.

Garrett, Henry E. Statistics in Psychology and Education. 4th ed. New York: Longmans, Green and Co., 1953.

Gatewood, Claude W., and Obourn, Ellsworth S. "Improving Science Education in the United States." Journal of Research in Science Teaching, 1 (1963), 355-399.

Gerberich, J. Raymond. Specimen Objective Test Items. New York: Longmans, Green and Co. 1956.

Good, Carter V., ed. Dictionary of Education. 2d ed. New York: McGraw-Hill Book Co., Inc., 1959.

Guilford, J.P. Fundamental Statistics in Psychology and Education. 3d ed. New York: McGraw-Hill Book Co., Inc., 1956.

_________. Psychometric Methods. New York: McGraw-Hill Book Co., Inc., 1954.

Guttman, Louis. "A Basis for Scaling Qualitative Data." American Sociological Review, 9 (April, 1944), 139-150.

_________. "Problems of Reliability." Measurement and Prediction, Vol. IV of Studies in Social Psychology in World War II. Edited by Samuel A. Stouffer, Princeton, N.J.: Princeton University Press, 1950. (Reissued as a Science Editions, New York: John Wiley and Sons, 1966.)

_________. "The Cornell Technique for Scale and Intensity Analysis." Educational and Psychological Measurement, (Summer, 1947), 247-279.

Haggard, Ernest A. Intraclass Correlation and the Analysis of Variance. New York: The Dryden Press Inc., 1958.

Haney, Richard E. The Changing Curriculum: Science. Washington, D.C.: Association for Supervision and Curriculum Development, National Education Association, 1966.

Harris, Chester W. "Review of This Issue." Review of Educational Research, 32, Chapter 9, (February, 1962)

Harris, Robin S. Quiet Evolution - A Study of the Educational System of Ontario. Toronto: University of Toronto Press, 1967.

Harvard Project Physics: Newsletter I. Harvard Project Physics, Harvard University, Fall, 1964.

Harvard Project Physics Progress Report. Harvard Project Physics, Harvard University, February, 1967.

Hedges, William D. Testing & Evaluation for the Sciences in the Secondary School. Belmont, Calif. Wadworth Publishing Co., Inc., 1966.

Herron, James Dudley. "A Factor Analytic and Statistical Comparison of CHEM Study and Conventional Chemistry in Terms of Their Development of Cognitive Abilities." Unpublished Ph.D. Dissertation. The Florida State University, Tallahassee, 1965. Dissertation Abstracts (1965): 4333.

________. "Evaluation and the New Curricula." Journal of Research in Science Teaching, 4 (1966): 159-170.

Holton, Gerald. "Modern Science and the Intellectual Tradition." Science, 131, (April 22, 1960), 1187-1192.

________. "Science for Nonscientists: Criteria for College Programs." Teachers College Record, 64 (1963): 497-509.

Hurd, Paul DeHart. "The New Curriculum Movement in Science: An Interpretive Summary." The Science Teacher, 29, (February, 1962), 6-9.

Hurd, Paul DeHart, et al. "Science Education for Changing Times." Rethinking Science Education. The Fifty-Ninth Yearbook of the National Society for the Study of Education, Part I, Chicago: The University of Chicago Press, 1960.

Johnson, Palmer O. Statistical Methods in Research. New York: Prentice-Hall, Inc., 1949.

Johnson, Philip G. "The Goals of Science Education." Theory Into Practice, 1 (December, 1962) 239-244.

Kaiser, Henry F. "Scaling a Simplex." Psychometrika, 27 (1962) 155-162.

Kellogg, Ralph E. American History Test Bank. Secondary Level. San Diego, Calif.: San Diego County Department of Education, 1964.

Kendall, Maurice G., and Stuart, Alan. The Advanced Theory of Statistics. Vol. II: Inference and Relationship. London: Charles Griffin & Company Limited, 1961.

Klinckmann, Evelyn. "The BSCS Grid for Test Analysis." BSCS Newsletter 19. Boulder, Colorado: The Biological Sciences Curriculum Study, 1963.

Klopfer, Leopold E., and Cooley, William W. "The History of Science Cases for High Schools in the Development of Student Understanding of Science and Scientists." Journal of Research in Science Teaching, 1 (1963), 33-47.

Klubertanz, George P. "The Nature of Science and the Teaching of High-School Chemistry." Journal of Chemical Education, 32 (1955), 248-252.

Krathwohl, David R., Bloom, Benjamin S., and Masia, Bertram B. Taxonomy of Educational Objectives. Handbook II: Affective Domain. New York: David McKay, Inc., 1964.

Kropp, Russell P., and Stoker, Howard W. The Construction and Validation of Tests of the Cognitive Processes as described in the Taxonomy of Educational Objectives. Cooperative Research Project No. 2117. Tallahassee, Fla.: Institute of Human Learning and Department of Educational Research and Testing, Florida State University, February, 1966.

Kropp, Russell P., Stoker, Howard W., and Bashaw, W. Louis. "The Validation of the Taxonomy of Educational Objectives." The Journal of Experimental Education, 34 (Spring, 1966), 69-76.

Lessinger, L. M. "Test Building and Test Banks Through the Use of the Taxonomy of Educational Objectives." California Journal of Educational Research, 14 (November, 1963), 195-201.

Lindvall, C.M., ed. Defining Educational Objectives. A Report of the Regional Commission on Educational Coordination and the Learning Research and Development Center. Pittsburgh: University of Pittsburgh Press, 1964.

Lockard, J. David. Sixth Report of the International Clearinghouse on Science and Mathematics Curricular Developments, 1968. Joint project of the American Association for the Advancement of Science, Washington, D.C., and Science Teaching Center, University of Maryland, College Park, Maryland, 1968.

Lombard, John W. "Preparing Better Classroom Tests." The Science Teacher, 32 (October, 1965), 33-38.

Lucow, William H. "Canadian Education." Science Education, 49 (October, 1965), 362-367.

Martin, W. Edgar. The Major Principles of the Biological Sciences of Importance for General Education. U.S. Office of Education Circular No. 308. Washington, D.C.: U.S. Government Printing Office, 1948.

Maykovich, John James. "Relationships Between Attitudinal Orientations and Academic Achievement in the Secondary School." Unpublished Ph.D. dissertation, University of California, Berkeley, 1966.

Mead, Margaret, and Métraux, Rhoda. "Image of the Scientist Among High School Students: A Pilot Study." Science, 126 (1957), 384-390.

Milholland, J.E. "Measuring Cognitive Abilities." Research on the Characteristics of Effective College Teaching. By W.J. McKeachie, R.L. Isaacson, and J.E. Milholland. Ann Arbor: University of Michigan, 1964.

Mills, Lester C., and Dean, Peter M. Problem-Solving Methods in Science Teaching. Science Manpower Project Monographs. New York: Bureau of Publications, Teachers College, Columbia University, 1960.

Modern High School Physics: A Recommended Course of Study. 2d ed. Science Manpower Project Monographs. New York: Bureau of Publications, Teachers College, Columbia University, 1959.

Monaghan, Floyd. "Design of Objective Test Items to Evaluate Thinking Ability in Science." Science Education, 44 (1960), 358-366.

Montean, John J., Cope, Ruth C., and Williams, Royce "An Evaluation of CBA Chemistry for High School Students." Science Education, 47 (1963), 35-43.

Morris, G. C. "Educational Objectives of Higher Secondary School Science." Unpublished M.Ed. Thesis, University of Sydney, Sydney, N.S.W., Australia, 1961.

Mursell, James L. Psychological Testing. 2d ed. New York: Longmans, Green & Co., 1949.

McFall, Robert W. "The Development and Validation of an Achievement Test for Measuring Higher Level Cognitive Processes in General Science." The Journal of Experimental Education, 33 (Fall, 1964), 103-106.

McGuire, Christine. "A Process Approach to the Construction and Analysis of Medical Examinations." The Journal of Medical Education, 38 (1963), 556-563. (a)

________. "Research in the Process Approach to the Construction and Analysis of Medical Examinations." National Council on Measurement in Education Yearbook, 20 (1963), 7-16. (b)

Nagel, Ernest. The Nature and Aim of Science. Philosophy of Science Series, No. 1, The Voice of America Forum Lectures. Washington, D.C.: U.S. Information Agency, 1963.

Nedelsky, Leo. Science Teaching and Testing. New York: Harcourt Brace and World, Inc., 1965.

Nelson, Clarence H. "Evaluation of Objectives of Science Teaching." Science Education, 43 (February, 1959), 20-27.

________. Let's Build Quality Into Our Science Tests. Washington, D.C.: National Science Teachers Association, 1958.

Novak, Joseph D. "A Preliminary Statement on Research in Science Education." Journal of Research in Science Teaching, 1 (1963), 3-9.

Nunnally, Jum C., Jr. Tests and Measurements. New York: McGraw-Hill Book Co., Inc., 1959.

Obourn, Ellsworth S. "Assumptions in Ninth-grade General Science." Unpublished doctoral dissertation, New York University, 1950.

Pella, Milton O. "Scientific Literacy and the H.S. Curriculum." School Science and Mathematics, 67 (1967) 346-356.

Pella, Milton O., O'Hearn, George T., and Gale, Calvin W. "Referents to Scientific Literacy." Journal of Research in Science Teaching, 4 (1966) 199-208. (a)

________. "Scientific Literacy - Its Referents." The Science Teacher, 33 (May, 1966), 44. (b)

Phillips, Charles E. The Development of Education in Canada. Toronto: W.J. Gage and Co., Ltd., 1957.

Pierce, Edward F. Modern High School Chemistry: A Recommended Course of Study. Science Manpower Project Monographs. New York: Bureau of Publications, Teachers College, Columbia University, 1960.

Popham, W. James. Educational Statistics - Use and Interpretation. New York: Harper & Row, 1967.

Porter, Marjorie Ruth, and Anderson, Kenneth E. "A Study of the Relationship of Specified Abilities in Chemistry to each other and Intelligence." Science Education, 43, (1959), 12-19.

Price, D.J. "The Exponential Curve of Science." Discovery, 17 (June, 1956), 240-243.

Problem Solving Through Science. San Francisco, Calif.: Northern California Science Committee, 1959.

Rosen, Albert. "Test - Retest Stability of MMPI Scales for a Psychiatric Population." Journal of Consulting Psychology, 17 (1953), 217-221.

Rust, Velma I. "Factor Analyses of Three Tests of Critical Thinking." Journal of Experimental Education, 29 (1960), 177-182.

Rutherford, F. James. "The Role of Inquiry in Science Teaching." Journal of Research in Science Teaching, 2 (1964), 80-84.

Rutledge, James A. "Changing Emphases in High School Science." Theory Into Practice, 1 (December, 1962), 264-270.

Scannell, Dale P., and Stellwagen, Walter R. "Teaching and Testing for Degrees of Understanding." California Journal of Instructional Improvement, 3 (1960), 8-14.

Schmadel, Elnora. "The Relationship of Creative Thinking Abilities to School Achievement." Unpublished Ed.D. Thesis, University of Southern California, 1960. Dissertation Abstracts, 21 (December, 1960), 1464-1465.

Schmitt, J.A., et al. "Cooperative Development of Locally Oriented Achievement Tests in Chemistry." Journal of Research in Science Teaching, 4 (1966), 85-91.

Science: An Interim Report of the Science Committee. Toronto: Ontario Curriculum Institute, 1963.

Science Education News. Washington, D.C.: American Association for the Advancement of Science, December, 1961.

Science in Secondary Schools. Ministry of Education, Pamphlet No. 38. London, England: Her Majesty's Stationery Office, 1960.

Sears, Paul B., and Kessen, William. "Statement of Purposes and Objectives of Science Education in School." Journal of Research in Science Teaching, 2 (1964), 3-6.

Secondary Modern Science Teaching, Part I. The Secondary Modern Schools Sub-Committee of the Science Masters' Association. London: John Murray, 1954.

Shamos, Morris H. "Science for Citizens." Saturday Review, 44 (September 16, 1961), 68-69.

________. "The Price of Scientific Literacy." National Association of Secondary School Principals Bulletin, 47 (April, 1963), 41-51.

Short, L. N. "How to Start a Revolution." Approach to Chemistry 1962. Edited by R.M. Gascoigne and F.S. Symes. Sydney, Australia: The University of New South Wales, 1962.

Siegel, Sidney. Non-Parametric Statistics: For the Behavioral Sciences. New York: McGraw-Hill Book Co., Inc., 1956.

Smith, Eugene R., Tyler, Ralph W., et al. Appraising and Recording Student Progress. Vol. III of Adventure in American Education. Progressive Education Association Commission on the Relation of School and College. New York: Harper and Brothers, 1942.

Smith, Richard B. "A Discussion of an Attempt at Constructing Reproducible Item Sets." Journal of Educational Measurement, 5 (1968), 55-60.

________. "An Analysis of the Scalability of the 'Knowledge' and 'Comprehension' Levels of the Taxonomy of Educational Objectives: Cognitive Domain." Paper read at the National Council on Measurement in Education, Chicago, February, 1965.

________, and Paterson, John. "A Measurement Problem in Action Research." Unpublished manuscript, Purdue University, 1965.

Sonquist, John A., and Morgan, James N. The Detection of Interaction Effects. A Report on a Computer Program for the Selection of Optimal Combinations of Explanatory Variables. Monograph No. 35. Ann Arbor, Mich.: Survey Research Center, Institute for Social Research, The University of Michigan, 1964.

Stanley, Julian C., and Bolton, Dale L. "Review of Taxonomy of Educational Objectives, The Classification of Educational Goals, Handbook I: Cognitive Domain, edited by Benjamin S. Bloom; and Specimen Objective Test Items, A Guide to Achievement Test Construction, by J. Raymond Gerberich." Educational and Psychological Measurement, 17 (1957), 631-634.

Stoker, H. W. and Kropp, R. P. "Measurement of Cognitive Processes." Journal of Educational Measurement, 1 (1964), 39-42.

Stouffer, Samuel A., et al. "A Technique for Improving Cumulative Scales." Public Opinion Quarterly, 16 (1952), 273-291. (Also in Sociological Studies in Scale Analysis, edited by Matilda White Riley, et al. New Brunswick, N.J.: Rutgers University Press, 1954.)

Strong, Lawrence E. "Ideas and Chemicals." School Science and Mathematics, 59 (1959), 167-170.

Strong, Lawrence E., and Benfey, O. Theodor. "Is Chemical Information Growing Exponentially?" Journal of Chemical Education, 37 (1960), 29-30.

Sutman, Frank X. "Mass Education and the New Science." Science Education, 50 (1960), 494-496.

Swineford, Frances. "Evaluation of Performance on Some Canadian University Students on College Entrance Examination Board Tests." The Ontario Institute for Studies in Education, Toronto, January - February, 1966. (Restricted Distribution).

Teaching for Critical Thinking in Chemistry. 3 reports of Summer Conferences for Science Teachers. Washington, D.C.: The National Science Teachers Association, 1958.

"The Reed College Conference on the Teaching of Chemistry." Journal of Chemical Education, 35 (1958), 54-55.

The Teaching of General Science. A Report by the General Science Sub-committee of the Science Masters' Association. London: John Murray, 1950.

The Teaching of Science in Secondary Schools, rev. ed. Joint Committee of the Incorporated Association of Assistant Masters and the Science Masters' Association. London: John Murray, 1958.

Thomas, Alice Miller. "Levels of Cognitive Behaviour Measured in a Controlled Teaching Situation." Unpublished Master's Thesis, Cornell University, 1965.

Thorndike, Robert L. The Concepts of Over- & Underachievement, New York: Bureau of Publications, Teachers College, Columbia University, 1963.

Thorndike, Robert L., and Hagen, Elizabeth. Measurement and Evaluation in Psychology and Education. 2d ed. New York: John Wiley and Sons, Inc., 1961.

Toops, Herbert A. "A Comparison, By Work-Limit and Time-Limit, of Item Analysis Indices for Practical Test Construction." Educational and Psychological Measurement, 20 (1960), 251-266.

Tyler, Louise L. and Okumu, Laura J. "A Beginning Step: a System for Analyzing Courses in Teacher Education." Journal of Teacher Education, 16 (1965), 438-444.

Tyler, Ralph W. "Permanence of Learning." Constructing Achievement Tests. Columbus: Ohio State University, 1934.

Urdal, Lloyd B. "Science Objectives and Evaluation for Tomorrow." The High School Journal, 43 (March, 1960) 313-319.

Uricheck, Michael J. "Research Proposal: An Attempt to Evaluate the Success of the CBA and CHEMS Chemistry Courses." Science Education, 51 (1967), 5-11.

Washton, Nathan S. Teaching Science Creatively in the Secondary Schools. Philadelphia: W.B. Saunders Co., 1967.

Watson, Fletcher G. "Toward Effective Research in Science Education." Theory Into Practice, 1 (1962), 277-283.

Watson, Goodwin, and Glaser, Edward M. Watson-Glaser Critical Thinking Appraisal. Forms AM and BM, 1952, Forms YM and ZM, 1964. Tarrytown, N.Y.: Harcourt, Brace and World.

White, J.H. Teaching Chemistry. London: University of London Press, 1957.

Winer, B.J. Statistical Principles in Experimental Design. New York: McGraw-Hill Book Co., Inc., 1962.

Winter, Stephen S., et al. A Study of Large Group - Small Group Instruction in Regents Chemistry Compared to Conventional Instruction. Buffalo: Education Research Center, School of Education, State University of New York at Buffalo, April, 1965.

Wise, Harold E. "A Determination of the Relative Importance of Principles of Physical Science for General Education." Science Education, 25 (1941), 371-379; 26 (1942), 8-12.

Zinn, Karl Lewis. "Validation of a Differential Test of Cognitive Objectives of the First Course in Psychology." Unpublished Ph.D. Thesis, University of Michigan, Ann Arbor, Mich., 1964. Dissertation Abstracts, 25 (1964), 3413-3414.